



Universidad
Carlos III de Madrid
www.uc3m.es

**ESCUELA POLITÉCNICA SUPERIOR
DEPARTAMENTO DE TELEMÁTICA**

TRABAJO FIN DE GRADO

Análisis y estudio de los servicios de Google como herramienta de marketing

Autor: Verónica Ruiz González

Titulación: Grado en Ingeniería Telemática

Profesor: Juan Miguel Carrascosa Amigo

Fecha: Leganés, Mayo de 2015

Agradecimientos

A todos los que me han acompañado
en este largo viaje.

*Hay una fuerza motriz más poderosa
que el vapor, la electricidad y la
energía atómica: la voluntad.*

Albert Einstein

Abstract

The Internet has naturally come into our lives, step by step, it has gained ground in the different areas of our daily habits. Nowadays, it's possible to do some online activities that they were impossible to imagine long time ago.

The countless benefits of the Internet are well-known by everyone. However, the use of the network involves some risks, such as the use of our personal data –that we share while we're navigating- by third parties.

In this line, the latest trends in online marketing focus on collecting such personal data to show the ad campaigns only to those who can be really interested in them.

The purpose of this document is to find out how companies use our personal data to customize those ad campaigns. So, we'll focus on the study and the analysis of the Google services as a marketing tool because talking about Google is talking about the Internet.

Resumen

Internet ha entrado en nuestras vidas de manera natural, poco a poco, ha ido ganando terreno en las distintas parcelas de nuestros hábitos cotidianos sin apenas percatarnos. Hoy en día, es posible realizar algunas actividades en línea que hace tiempo simplemente eran imposibles de imaginar.

Los incontables beneficios de Internet son bien conocidos por todos. Sin embargo, la utilización de la red también entraña algunos riesgos, como el uso que hacen terceros de la información personal que compartimos al navegar.

En esta línea, las tendencias actuales del marketing online se centran precisamente en extraer esa información personal para mostrar las diferentes campañas de anuncios solo a aquellas personas que pudieran estar realmente interesadas en ellas.

En este trabajo, intentaremos descubrir cuál es realmente el uso que las distintas empresas dan a nuestra información personal para personalizar dichos anuncios. Para ello, nos centraremos en el estudio y análisis de los servicios de Google como herramienta de marketing, ya que hablar de Google es hablar de Internet.

ÍNDICE DE CONTENIDOS

Agradecimientos	2
Abstract	3
Resumen	4
ÍNDICE DE CONTENIDOS	5
ÍNDICE DE TABLAS	7
ÍNDICE DE FIGURAS	8
1. MOTIVATION AND OBJECTIVES	10
1.1 Motivation	10
1.2 Delineation of the Problem	12
1.3 Objectives	13
1.4 Structure of the Document	14
1. MOTIVACIÓN Y OBJETIVOS	15
1.1 Motivación	17
1.2 Delimitación del Problema	17
1.3 Objetivos	18
1.4 Estructura del Documento	19
2. PLANTEAMIENTO DEL PROBLEMA	21
2.1 Estado del Arte: Marketing en Internet	21
2.1.1 Publicidad en Internet	22
2.1.2 Posicionamiento en Buscadores	23
2.1.2.1 Posicionamiento SEO	24
2.1.2.2 Posicionamiento SEM	25
2.1.3 Marketing en Redes Sociales	25
2.1.3.1 Marketing en Facebook	26
2.1.3.2 Marketing en Twitter	27
2.1.3.3 Marketing en LinkedIn	28
2.2 Marco Regulador: Leyes de Protección de datos	28
2.2.1 Leyes Españolas	29
2.2.2 Leyes Europeas	30
2.3 Especificación del Problema: Google	30
2.3.1 Servicios de Google	32
2.3.1.1 Buscador	32
2.3.1.2 Gmail	34
2.3.1.3 Google Maps	35
2.3.1.4 Youtube	36
2.3.2 Servicios de publicidad de Google	37
2.3.2.1 Google Adwords	37
2.3.2.2 Google Adsense	37
2.3.3 Configuración de Anuncios de Google	38
3. DISEÑO DE LA SOLUCIÓN TÉCNICA	40
3.1 Análisis de Requisitos	40
3.2 Diseño de la Solución Técnica	46
3.2.1 Arquitectura	46
3.2.2 Tecnologías	49

3.2.2.1 Java	49
3.2.2.2 Selenium	51
3.2.2.3 MySQL	53
3.3 Implementación de la Solución Técnica	54
3.3.1 Diagrama de Clases	54
3.3.2 Modelo	55
3.3.2.1 Diseño Base de Datos	55
3.3.2.1.1 Modelo Entidad-Relación	55
3.3.2.1.2 Descripción de Tablas	56
3.3.3 Controlador	57
3.3.3.1 Gestión Base de Datos	58
3.3.3.2 Gestión Pruebas y Simulación Usuario	60
3.3.3.3 Extracción Inteligente	64
3.3.4 Vista	65
3.3.5 Características de la Herramienta	70
4. RESULTADOS Y EVALUACIÓN	72
4.1 Descripción de la Evaluación	72
4.2 Resultados en el Buscador	73
4.3 Resultados en Google Maps	75
4.4 Resultados en Youtube	76
4.5 Resultados en Gmail	79
4.6 Comparativa entre Servicios	79
4.7 Resultados Categorías Sensibles	80
5. PRESUPUESTO Y PLANIFICACIÓN DEL TRABAJO	82
5.1 Planificación del Trabajo	82
5.1.1 Fases del Ciclo de Vida	82
5.1.2 Diagrama de Gantt	83
5.2 Presupuesto del Trabajo	85
6. CONCLUSIONS	87
6. CONCLUSIONES	88
GLOSARIO DE TÉRMINOS	89
GLOSARIO DE ACRÓNIMOS	90
REFERENCIAS	91
APPENDIX A - EXTENDED SUMMARY	92

ÍNDICE DE TABLAS

Tabla 3.1: RQ-001: Automatización del Buscador	40
Tabla 3.2: RQ-002: Automatización de Gmail	40
Tabla 3.3: RQ-003: Automatización de Google Maps.	40
Tabla 3.4: RQ-004: Automatización de Youtube	40
Tabla 3.5: RQ-005: Automatización de Login	41
Tabla 3.6: RQ-006: Automatización de Logout	41
Tabla 3.7: RQ-007: Extracción Inteligente	41
Tabla 3.8: RQ-008: Almacenamiento General	41
Tabla 3.9: RQ-009: Almacenamiento Individual.	41
Tabla 3.10: RQ-010: Personalización de la Prueba	42
Tabla 3.11: RQ-011: Selección de Servicio	42
Tabla 3.12: RQ-012: Personalización de Correo Electrónico	42
Tabla 3.13: RQ-013: Selección de Categorías	42
Tabla 3.14: RQ-014: Gestión de Ficheros	43
Tabla 3.15: RQ-015: Selección de Palabra	43
Tabla 3.16: RQ-016: Selección de Contenido.	43
Tabla 3.17: RQ-017: Interacción con los Servicios.	43
Tabla 3.18: RQ-018: Rastreo	43
Tabla 3.19: RQ-019: Validación de Datos Introducidos	44
Tabla 3.20: RQ-020: Sucesión de Pruebas	44
Tabla 3.21: RQ-021: Gestión de Errores.	44
Tabla 3.22: RQ-022: Interfaz de Usuario Intuitiva.	44
Tabla 3.23: RQ-023: Selección de Opciones en la Interfaz de Usuario	44
Tabla 3.24: RQ-024: Valores por Defecto en la Interfaz de Usuario	45
Tabla 3.25: RQ-025: Botón de Información en la Interfaz de Usuario	45
Tabla 3.26: RQ-026: Velocidad del Sistema	45
Tabla 3.27: RQ-027: Rendimiento del Sistema	45
Tabla 3.28: RQ-028: Seguridad del Sistema	45
Tabla 3.29: RQ-029: Portabilidad del Sistema.	46
Tabla 3.30: RQ-030: Escalabilidad del Sistema.	46
Tabla 3.31: Tabla categorias	56
Tabla 3.32: Tabla pruebas_general	56
Tabla 3.33: Tabla pruebas_individual_google	57
Tabla 3.34: Tabla pruebas_individual_web	57
Tabla 4.1. Configuración de Pruebas	72
Tabla 4.2: Resultados en el Buscador	73
Tabla 4.3: Resultados en Google Maps	75
Tabla 4.4: Resultados en Youtube.	76
Tabla 4.5: Resultados Categorías Sensibles.	80
Tabla 5.1: Subtotal Personal.	85
Tabla 5.2: Subtotal Material.	85
Tabla 5.3: Presupuesto Total	86

ÍNDICE DE FIGURAS

Figure 1.1: Internet. Source: plus.google.com	10
Figure 1.2: Lack of privacy on the Internet. Source: www.neoteo.com	12
Figura 1.1*: Internet. Fuente: plus.google.com	15
Figura 1.2*: Falta de privacidad en Internet. Fuente: www.neoteo.com	17
Figura 2.1: Marketing en Internet. Fuente: plus.google.com	21
Figura 2.2: Anuncio de texto en Google	22
Figura 2.3: Banner animado en www.elmundo.es	22
Figura 2.4: Anuncio de vídeo en Youtube	23
Figura 2.5: Posicionamiento SEO y SEM en Google. Fuente: bacaklik.net	23
Figura 2.6: Cuota de mercado de Google en España. Fuente: Stat Counter.	24
Figura 2.7: Redes sociales más utilizadas en España. Fuente: iab	26
Figura 2.8: Anuncio en Facebook	27
Figura 2.9: Tweet promocionado	27
Figura 2.10: Anuncio en LinkedIn	28
Figura 2.11: Página principal de Google, cuyo Doodle (logotipo) es cambiante, según diversos eventos relevantes	33
Figura 2.12: Resumen del historial de búsquedas de Google	33
Figura 2.13: Tendencias del historial de búsquedas de Google	34
Figura 2.14: Anuncios de Google Adwords en Gmail	34
Figura 2.15: Resumen del historial de Gmail	35
Figura 2.16: Anuncios de Google Adwords en Google Maps.	35
Figura 2.17: Resumen del historial de búsquedas y del historial de reproducciones en Youtube	36
Figura 2.18: Anuncios de Google Adsense	38
Figura 2.19: Configuración de anuncios de Google	39
Figura 3.1: Arquitectura del Sistema	48
Figura 3.2: Ejemplo de JavaFX. Fuente: tutorialjavafx.blogspot.com.es	50
Figura 3.3: Interfaz de Selenium IDE. Fuente: docs.seleniumhq.org	52
Figura 3.4: Diagrama de Clases.	54
Figura 3.5: Modelo Entidad-Relación de la base de datos	55
Figura 3.6: Clase BDAnuncios.	58
Figura 3.7: Clase GoogleAds.	60
Figura 3.8: Diagrama de flujo de la gestión de pruebas. Método comenzar	61
Figura 3.9: Diagrama de flujo de la automatización del Buscador, de Google Maps y de Youtube	63
Figura 3.10: Diagrama de flujo de la automatización de Gmail.	63
Figura 3.11: Clase Analisis	64
Figura 3.12: Clase Ventana.	65
Figura 3.13: Interfaz Gráfica de Usuario	66
Figura 3.14: Selección Fichero de Entrada en la Interfaz Gráfica de Usuario	67
Figura 3.15: Error en la validación de los datos introducidos en la Interfaz Gráfica de Usuario.	68
Figura 3.16: Incidencia ocurrida durante la prueba.	69
Figura 3.17: Prueba Finalizada con Éxito.	69

Figura 3.18: Manual de Usuario en la Interfaz Gráfica de Usuario	70
Figura 4.1: Resultados en Gmail.	79
Figura 4.2. Gráfico comparativo en número de palabras que generan intereses. 79	
Figura 4.3. Gráfico comparativo en tiempo medio de categorización	80
Figura 4.4. Gráfico comparativo en relación semántica	80
Figura 4.5. Gráfico comparativo entre set 1 y set 2.	81
Figura 5.1: Modelo Cascada Realimentado. Fuente: es.wikipedia.org	82
Figura 5.2: Tareas Realizadas	83
Figura 5.3: Diagrama de Gantt.	84
Figure A.1: Settings for Google Ads.	94
Figure A.2: Graphical User Interface.	95
Figure A.3: System Architecture	96
Figure A.4 Number of words generating interests	98
Figure A.5: Average sorting time	98
Figure A.6: Semantic Relationship.	98
Figure A.7: Set 1 vs Set 2	99

1. MOTIVATION AND OBJECTIVES

1.1 Motivation

The unstoppable advance of **the Internet has totally changed our habits**, it has been a full revolution because it has affected all areas of life.

Some important areas are: communication and interaction with others (e-mail, blogs, social networks, instant messaging), shopping (e-commerce, e-money), bank accounts management, business (e-business, cloud storage). Also how to read and write, learning (e-learning, MOOC), job search (job sites), information (online newspapers, specialized websites, wikis), listening to music, watching videos, planning trips, etc... In general, the manner to enjoy leisure and life, and so on, to complete the most common areas of anyone's life.

It's not possible to imagine life without the Internet, who has not an email account from Gmail, Outlook or Yahoo Mail? Who has never read, or written, a Blogger or Wordpress blog? Who isn't member of famous social networks like Facebook, Twitter or LinkedIn? Who has never used Skype or WhatsApp? Who has never visited the eBay or Amazon sites for shopping? Who has never paid through PayPal on the network? Who has never stored files on Dropbox or Google Drive? Who has never listened to music on Spotify? And who has never watched a video on Youtube?



Figure 1.1: Internet. Source: plus.google.com

It's difficult to answer negatively any last questions. The medium has changed and the users have changed too. Moreover, due to the boom of mobile devices, it's possible to be continuously connected, placing the Internet as the ideal place to find new customers for companies.

Traditional marketing has become completely outdated and companies have had to adapt to the new situation -where millions of people from around the world daily use the Internet-, creating new ways to attract the attention of potential clients. And this new kind of marketing is known as **online marketing**.

The Internet is quite different from previous media, it's interactive and it allows to get feedback from customers, also to build a closer relationship with them: to know their preferences and their online shopping trends. And so, to be able to customize online marketing strategies to potential customers.

It's unsuccessful to show all ad campaigns to all users. Moreover, the network features help to the content customization, and so, the ad campaigns can be shown only to those users who can be really interested in them. In this line, the latest trends in advertising focus on the audience **segmentation according to the user profiles** on the network.

These techniques -used by advertisers and website administrators- collect information from the web browsers: web history, search history, amount of time on each page, the visited links, among others. Another factors are: geography, demographics or contextual content of the site. In this way, it's possible to create an user profile when anyone navigates on the Internet.

But, are we really aware of all information that we share only by clicking?

Web navigation is a voluntary activity -although not so much-. However, it's not an anonymous activity -as anyone could think-. Both Internet providers and website administrators can collect a large amount of personal data during navigation (for example, using cookies).

Due to the uncertain privacy that we have while accessing the Internet, is the **privacy** of our **personal data** insured on the network?

Nowadays the need for using the Internet -by most of the world's population- has led the implantation of some international laws on **data protection**, that try to protect the users' privacy as much as possible, especially the data that are related to the **sensitive categories** -those that need special attention for being linked to the personal privacy, fundamental rights and civil liberties-.



Figure 1.2: Lack of privacy on the Internet. Source: www.neoteo.com

After reading the above information, some reasonable doubts could appear regarding the methods of customization -that companies use to develop their online marketing strategies-. Specifically, if the methods infringe the users' privacy on the Internet and, especially, if they infringe the processing of sensitive categories. Because of the users aren't really aware of sharing so much information only by clicking on a link or only by creating an account on a site.

1.2 Delineation of the Problem

This work is going to focus on studying and analyzing **Google** as a marketing tool: how our personal data are collected -using some basic services (**Search engine, Gmail, Google Maps and Youtube**)- to place interest-based ads inside its advertising network and how the sensitive categories are processed.

According to the information that Google provides, its network coverage -including specific Google sites, applications, mobile sites and millions of websites: news pages, blogs, etc...- reaches 90 % of Internet users. This high percentage places Google as the largest collector of personal data around the world, and this is the reason why it's going to be the focus of this study.

1.3 Objectives

After introducing the motivation, that has originated this study, and also after specifying the problem to solve -Google services as a marketing tool-. The objectives to achieve are the followings:

- Identify and analyze the main online marketing strategies.
- Identify the most important laws on data protection in Spain and Europe. Specifically, those that are related to the processing of **sensitive categories** on the network.
- Study and analyze the basic **Google services**, its advertising network and its privacy policy.
- Design and develop a flexible tool to **simulate** the **user behavior** and so to automatically use the Google services.
- Study the Selenium technology as a browser automation tool.
- Develop a tool to perform an **intelligent data extraction** from the user preferences in settings for Google Ads.
- Design and manage a database to store data extracted.
- Design and develop a graphical user interface according to the basic design criteria.
- Design and perform a test plan, inside the **evaluation**, to determine how the interest-based ads work.
- Evaluate the results.
- Evaluate the processing of sensitive categories by Google.
- Estimate a budget and a project planning.
- Draw conclusions and make future research suggestions.

1.4 Structure of the Document

This section is dedicated to explain how the document is structured, and so, to help the understanding of the reader:

- **Abstract:** A brief summary of the work, that discloses the main issues to be discussed in it. Firstly, the English version, secondly, the Spanish one.
- **Table of contents and index of tables and figures.**
- **Chapter 1 - Motivation and Objectives:** It introduces the motivation of the work and the specific problem to solve. It also contains the definition of the objectives to achieve. Firstly, the English version, secondly, the Spanish one.
- **Chapter 2 - Problem Statement:** It includes the vision of the state of the art -the current situation of the online marketing-, the identification of the regulatory framework -laws on data protection- and the description of the problem -Google, its services and its privacy policy-.
- **Chapter 3 - Technical Design Solution:** It contains the requirements analysis, the description of system architecture, the technologies to use and the specification of the different modules of the architecture.
- **Chapter 4 - Results and Evaluation:** It includes the description of the evaluation, the results for each service, the comparison between them and the evaluation of the processing of sensitive categories.
- **Chapter 5 - Budget and Planning:** It contains the project planning for the different phases of life cycle and the estimation of the budget.
- **Chapter 6 - Conclusions:** It includes the conclusions of the work and the future research suggestions. Firstly, the English version, secondly, the Spanish one.
- **Glossary of terms and acronyms.**
- **References.**
- **Appendix A - Extended Summary:** It contains the description of the problem and its context, the technical solution and the results of the evaluation. English version.

1. MOTIVACIÓN Y OBJETIVOS

1.1 Motivación

El avance imparable de **Internet** ha **cambiado drásticamente** **nuestros hábitos**, ha supuesto una revolución integral, ya que ha afectado a todos los ámbitos de la vida.

Desde la forma de comunicarnos, de relacionarnos (correo electrónico, blogs, redes sociales, mensajería instantánea), de comprar (e-commerce, dinero electrónico), de gestionar nuestras cuentas bancarias, de trabajar (e-business, sistemas de almacenamiento en la nube), pasando por la forma de leer y escribir, de aprender (e-learning, MOOC), de buscar trabajo (portales de empleo), de informarnos (periódicos digitales, webs especializadas, wikis), de escuchar música, de visualizar vídeos, de planificar nuestros viajes, etc... En general, de disfrutar del ocio y de la vida, y así sucesivamente, hasta albergar el resto de áreas cotidianas de cualquier persona.

Resulta inviable imaginarse la vida sin Internet, ¿quién no tiene una cuenta de correo electrónico de Gmail, de Outlook o de Yahoo Mail? ¿Quién no ha leído, o escrito, alguna vez un blog de Blogger o de Wordpress? ¿Quién no es usuario de redes sociales tan conocidas como Facebook, Twitter o LinkedIn? ¿Quién no se ha comunicado alguna vez por Skype o por Whatsapp? ¿Quién no ha visitado alguna vez los portales de eBay o de Amazon para realizar sus compras? ¿Quién no ha pagado alguna vez a través de PayPal en la red? ¿Quién no ha almacenado alguna vez sus archivos en Dropbox o en Google Drive? ¿Quién no ha escuchado alguna vez música con Spotify? ¿Y quién no ha visto alguna vez un vídeo en Youtube?



Figura 1.1*: Internet. Fuente: plus.google.com

Es prácticamente imposible responder negativamente a cualquiera de las preguntas formuladas en el párrafo anterior. El medio ha cambiado y los usuarios hemos cambiado con él. Además gracias al auge de los dispositivos móviles, es posible estar conectado continuamente, situando la red como el lugar ideal para que las empresas capten nuevos clientes.

El marketing tradicional se ha quedado completamente obsoleto y las empresas se han tenido que adaptar a la nueva situación -donde millones de personas de todo el mundo se encuentran a diario en Internet-, creando nuevas fórmulas para llamar la atención de los posibles consumidores. Y este nuevo enfoque del marketing es lo que se conoce como **marketing en Internet**.

Internet es un medio de comunicación completamente diferente a todos los anteriores, es interactivo y permite tener retroalimentación por parte de los consumidores, así como entablar una relación más estrecha con ellos: conocer sus preferencias, sus tendencias de consumo. Y de esta manera, personalizar las estrategias de marketing online a los posibles clientes.

Mostrar todas las campañas publicitarias a todos los usuarios es poco efectivo. Especialmente, si las características de la red facilitan la personalización de contenidos y mostrar dichas campañas solo a los usuarios que realmente puedan estar interesados en ellas. Por esta razón, la tendencia actual de la publicidad se centra en la **segmentación** de las campañas **según** el **perfil** de los usuarios en la red.

Estas técnicas -utilizadas por los publicistas y por los administradores de páginas web para segmentar los anuncios- recopilan información de los navegadores web de los usuarios: como el historial web, el historial de búsquedas, el tiempo de navegación en cada página, los enlaces en los que se hace clic, entre otros. También se tienen en cuenta factores como la geografía, la demografía o el contenido contextual de la página. De esta manera, se puede crear un perfil del tipo de usuario al que corresponde cuando navega por Internet.

Pero, ¿realmente somos conscientes de toda la información que compartimos solo con hacer un clic?

La navegación es una actividad voluntaria -aunque ya no tanto-. Sin embargo, no es una actividad anónima -como podría pensarse-. Tanto los proveedores de Internet como los administradores de páginas web tienen la capacidad de recopilar una gran cantidad de información personal mientras navegamos por la red (por ejemplo, con el uso de cookies).

Debido a la dudosa privacidad con la que contamos al acceder a la red, ¿hasta qué punto se garantiza la **privacidad** de nuestros **datos personales**?

En la actualidad, la necesidad del uso de Internet -para la mayor parte de la población mundial- ha provocado la implantación de una serie de leyes internacionales de

protección de datos, para intentar proteger la privacidad de los internautas en la medida de lo posible, sobre todo, con respecto a los datos de **categorías sensibles** -aquellos que requieren una mayor privacidad al estar ligados con la intimidad de las personas, sus derechos fundamentales y sus libertades públicas-.



Figura 1.2*: Falta de privacidad en Internet. Fuente: www.neoteo.com

Como se puede comprobar con la información expuesta hasta ahora, pueden surgir dudas bastante razonables sobre si los métodos de personalización de anuncios -que utilizan las empresas para desarrollar sus nuevas estrategias de marketing- vulneran la privacidad de los usuarios al navegar por la red, especialmente con respecto al trato que debe darse a los datos de categorías sensibles. En muchas ocasiones, sin ser los usuarios realmente conscientes de la información que comparten, solo por hacer un clic en una página o solo por crearse una cuenta en un sitio determinado.

1.2 Delimitación del Problema

Este trabajo se va a centrar en estudiar y analizar al gigante **Google** como herramienta de marketing: la manera en que extrae nuestra información personal y nuestras preferencias -a la hora de usar sus servicios más conocidos (**Buscador, Gmail, Google Maps y Youtube**)- para segmentar por intereses los anuncios de su compleja red de anunciantes y el trato que da a nuestros datos de categorías sensibles.

Según la información que el propio Google nos facilita, la cobertura de su red -que alberga sitios web específicos de Google, aplicaciones, sitios para móviles, así como millones de sitios web, páginas de noticias, blogs, etc...- llega al 90% de los usuarios de Internet de todo el mundo. Este elevado porcentaje sitúa a Google como el mayor recolector de información personal del mundo, razón por la que va a ser el centro de este estudio.

1.3 Objetivos

Una vez expuestas las motivaciones que llevan a la realización de este estudio, y delimitado el problema al caso concreto de los servicios de Google como herramienta de marketing. Los objetivos a cumplir serán los siguientes:

- Determinar y analizar las principales estrategias de marketing online en la actualidad.
- Identificar las leyes más relevantes de protección de datos que rigen en España y en Europa. Concretamente, determinar el trato que debe darse a los datos de **categorías sensibles** de los individuos en la red.
- Estudiar y analizar los **servicios** de **Google** más conocidos, sus servicios de marketing y su política de privacidad.
- Diseñar y elaborar una herramienta flexible, que permita la **simulación** del **comportamiento** de un **usuario**, para utilizar los servicios de Google de manera automática.
- Estudiar la tecnología Selenium como herramienta para automatizar la interacción con sitios web.
- Desarrollar una herramienta capaz de hacer una **extracción inteligente** de los datos relacionados con las preferencias de un usuario, respecto a la configuración de los anuncios segmentados por intereses de Google.
- Diseñar y gestionar una base de datos con los datos extraídos de la web.
- Diseñar y desarrollar una interfaz gráfica de usuario que cumpla con los criterios básicos de diseño.
- Diseñar y ejecutar un plan de pruebas, como parte de la **evaluación**, para determinar el funcionamiento de los anuncios basados en intereses de Google.
- Evaluar los resultados obtenidos.
- Evaluar el trato dado a los datos de categorías sensibles por parte de Google.
- Elaborar un presupuesto y una planificación para el proyecto.
- Desarrollar conclusiones y líneas futuras de investigación.

1.4 Estructura del Documento

A continuación, se va a explicar cómo se ha organizado el presente documento para facilitar la comprensión del lector:

- **Resumen:** Breve resumen del trabajo, en el que se dan a conocer los principales aspectos que serán tratados en el mismo. Primero, en su versión inglesa, y posteriormente, en su versión en castellano.
- **Índices:** Corresponde tanto con el índice general de contenidos como con los índices de figuras y tablas.
- **Capítulo 1 – Motivación y Objetivos:** Presenta la motivación que lleva a realizar el estudio, delimita el problema a tratar y define los objetivos que se pretenden alcanzar con el mismo. Primero, en su versión inglesa, y posteriormente, en su versión en castellano.
- **Capítulo 2 – Planteamiento del Problema:** Primero, analiza el estado del arte: la situación actual del marketing en Internet. A continuación, describe el marco regulador: las leyes de protección de datos que se encuentran en vigor, así como el trato que debe darse a los datos de categorías sensibles. Y finalmente, se centra en el problema a tratar: Google, sus servicios y su política de privacidad.
- **Capítulo 3 – Diseño de la Solución Técnica:** Comienza con el análisis de requisitos del problema. Posteriormente, detalla la arquitectura de la solución elegida, así como las tecnologías a emplear en la misma. Finalmente, describe la implementación de los distintos módulos de la arquitectura elegida.
- **Capítulo 4 – Resultados y Evaluación:** Primero, describe la evaluación a realizar. A continuación, evalúa los resultados obtenidos y muestra una serie de gráficas comparativas entre los diferentes servicios estudiados. Finalmente, evalúa el resultado del trato que Google da a los datos de categorías sensibles.
- **Capítulo 5 – Presupuesto y Planificación del Trabajo:** Comienza con la planificación de las distintas fases del proyecto. A continuación, detalla la estimación del presupuesto de los medios empleados para llevar a cabo el estudio.
- **Capítulo 6 – Conclusiones:** Describe las conclusiones obtenidas tras la realización del trabajo. También propone futuras líneas de investigación para continuar con el estudio del tema tratado. Primero, en su versión inglesa, y posteriormente, en su versión en castellano.
- **Glosario de términos y de acrónimos** que aparecen en el documento.
- **Referencias:** Se incluyen todos los recursos empleados en la elaboración del trabajo.

- **Apéndice A – Resumen Extendido** : Breve descripción del problema y su contexto, de la solución elegida y de los resultados obtenidos en la evaluación. Escrito en lengua inglesa.

2. PLANTEAMIENTO DEL PROBLEMA

2.1 Estado del Arte: Marketing en Internet

El marketing en Internet (digital, online, 2.0 o e-marketing) es la transformación sufrida por el marketing tradicional a causa del avance de Internet en la sociedad. Su filosofía se basa en que el público sea el centro de las campañas publicitarias y es necesario que exista una interacción entre las partes. Por lo tanto, se abandona la creatividad en favor de los contenidos.

El objetivo principal continua siendo promocionar y vender productos y servicios. Sin embargo, las estrategias empleadas para tal fin han cambiado radicalmente. Entre las nuevas tareas a realizar pueden incluirse: la gestión de contenidos, las relaciones públicas, la reputación en línea, el servicio al cliente y las ventas.



Figura 2.1: Marketing en Internet. Fuente: plus.google.com

Las estrategias más conocidas de marketing en Internet son: diseño web, anuncios en páginas web (banners u otros formatos publicitarios), posicionamiento en buscadores (tanto orgánico -SEO- como pago por clics -SEM-), envío de correos masivos (e-mail marketing), la utilización de las redes sociales (Social Media Marketing), así como de las redes de afiliados.

Otras estrategias son: marketing móvil, marketing viral, marketing de blogs, marketing de microblogs, podcasting, vídeo marketing, widgets o wikimarketing. Además están las campañas de resultados (Performance Marketing) -con costes por clic, por impresión, por cliente potencial o por venta- y la auditoría web -para medir la reputación, la popularidad y la influencia de una determinada página web y así poder orientar las estrategias publicitarias-.

En esta sección, analizaremos las estrategias de marketing online más relevantes para nuestro problema.

2.1.1 Publicidad en Internet

Existen numerosos formatos para mostrar publicidad online: texto, enlace, banner, web, weblog, blog, logo, anuncio, audio, vídeo y animación. También formatos más modernos como: videojuego, mensajería instantánea, descarga, interacción con SMS, así como otros formatos interactivos.

A continuación, se describen los formatos más significativos para el problema en cuestión:

- **Texto:** Consiste en una breve descripción del producto promocionado, además intenta captar la atención del posible consumidor. Los anuncios de texto suelen contar con un título, un eslogan, una URL y un enlace a dicha URL.

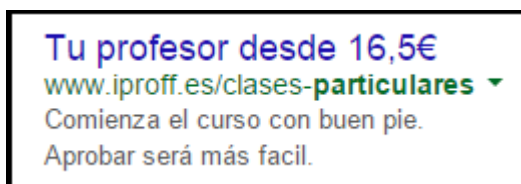


Figura 2.2: Anuncio de texto en Google

- **Banner:** Consiste en un texto, una imagen o una animación que al hacer clic enlaza con la página web del producto promocionado. Puede ser estático, animado o rotativo. Las animaciones suelen realizarse con flash.



Figura 2.3: Banner animado en www.elmundo.es

- **Vídeo:** Consiste en animaciones de formato multimedia. Puede contener texto, audio, imágenes y fotografías. Se realiza con flash o se graba directamente con cámaras digitales o de vídeo. Suele durar de 30 segundos a 2 minutos. Además, puede provocar campañas de marketing viral si capta el interés del usuario.



Figura 2.4: Anuncio de vídeo en Youtube

2.1.2 Posicionamiento en Buscadores

Existen dos tipos de posicionamiento en buscadores: **SEO** y **SEM**. Cuya finalidad es optimizar la visibilidad de los sitios web en los motores de búsqueda, con la utilización de palabras clave.

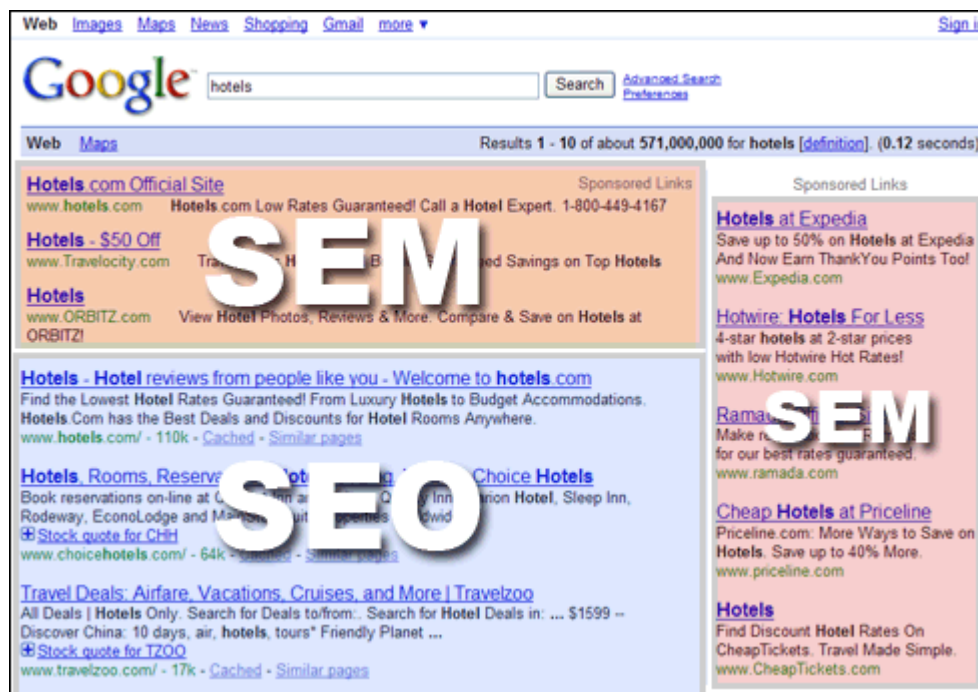


Figura 2.5: Posicionamiento SEO y SEM en Google. Fuente: bacaklik.net

2.1.2.1 Posicionamiento SEO

El posicionamiento SEO (Search Engine Optimization) consiste en la utilización de técnicas adecuadas -por los administradores de páginas web y proveedores de contenido- para que determinadas páginas web aparezcan -de manera natural- entre los primeros resultados de una búsqueda, al utilizar una serie de palabras clave.

El posicionamiento natural depende -principalmente- de la estructura y del contenido del sitio web: título, metas, palabras clave, accesibilidad, entre otros. Esto es lo que se conoce como optimización interna (On-page SEO).

La optimización externa (Off-page SEO) consiste en una serie de técnicas específicas para mejorar la notoriedad de la web como: linkbuilding y linkbaiting (conseguir que la página sea enlazada desde otros sitios web para hacerla más relevante), así como la creación de contenidos virales y la aplicación de estrategias correctas en las redes sociales.

El posicionamiento SEO se realiza mayoritariamente en **Google**, cuya cuota de mercado alcanza el 90-95% del volumen total de búsquedas en España (Figura 2.6). A su vez, la multinacional aplica una serie de penalizaciones a los sitios web que abusan de las manipulaciones para posicionarse lo más alto posible en los resultados de sus búsquedas. La última actualización del algoritmo de su buscador (Panda 4.0) fue lanzada el 20 de Mayo de 2014.

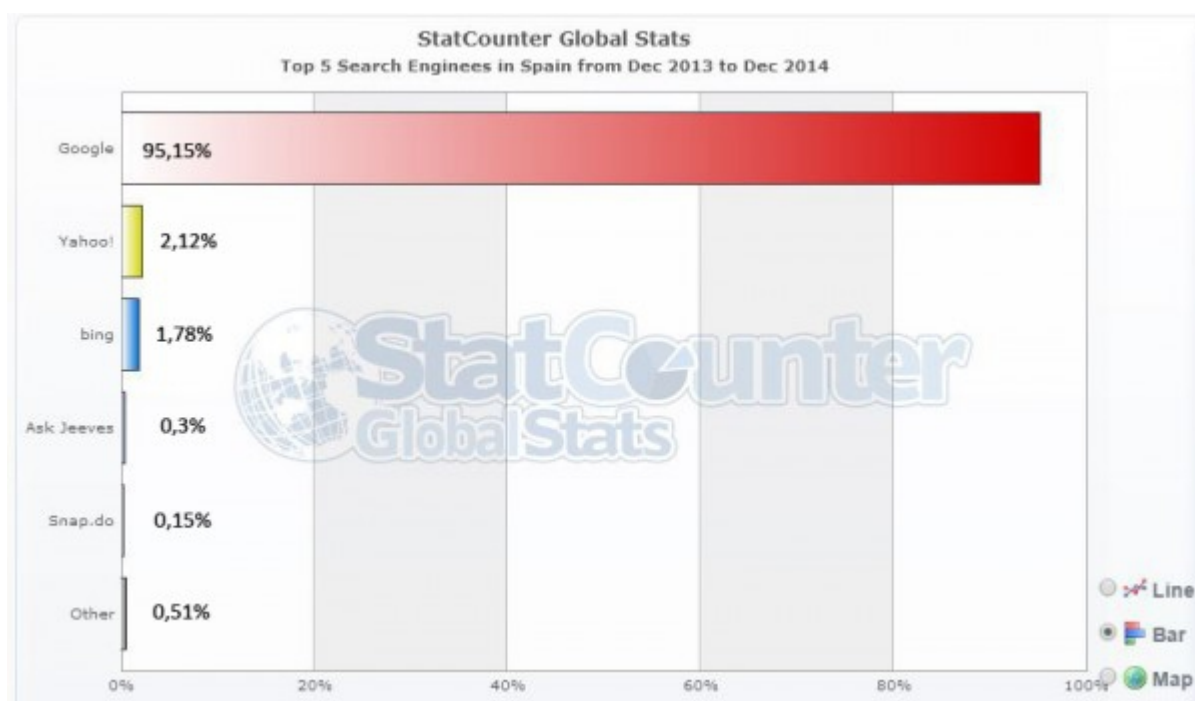


Figura 2.6: Cuota de mercado de Google en España. Fuente: Stat Counter.

Los principales buscadores (Google, Yahoo! y Bing) facilitan herramientas -a los administradores de sitios web- para tratar de optimizar el posicionamiento SEO. Entre ellas, Google ofrece: Google Analytics y la herramienta de webmaster de Google.

2.1.2.2 Posicionamiento SEM

El posicionamiento SEM (Search Engine Marketing) consiste en la contratación de campañas de marketing para que el sitio web aparezca entre los resultados destacados, cuando se realiza la búsqueda de determinadas palabras clave. También se le conoce como pago por clic -Pay Per Click (PPC)-.

Con el posicionamiento pagado se consigue: aumentar el tráfico de una web, dar a conocer una marca en un segmento de mercado determinado, tener mayor notoriedad de manera económica, dirigir anuncios hacia sectores que nos interesen y crear campañas de publicidad contextual personalizada.

Los principales buscadores (Figura 2.6) nos ofrecen una serie de productos para crear campañas publicitarias a través de sus motores de búsqueda.

Google nos ofrece **Google AdSense** y **Google Adwords**. El primero para crear campañas de publicidad contextual en sitios web de particulares. Y el segundo, para crear campañas de publicidad de búsqueda en su red de display. Más adelante se comentarán en detalle.

Por su parte, Bing y Yahoo! nos ofrecen -conjuntamente- su producto Bing Ads, para crear campañas de anuncios segmentadas, que sean mostradas durante la utilización de dichos motores de búsqueda. También permiten la posibilidad de crear campañas de publicidad contextual con Yahoo! Bing Network Contextual Ads.

2.1.3 Marketing en Redes Sociales

Las redes sociales fueron diseñadas como lugar de encuentro virtual, donde millones de personas pudieran compartir diversos intereses comunes. Tal ha sido la aceptación entre los internautas, que se han convertido en uno de los mejores lugares para mostrar campañas de marketing online.

Existen diferentes tipos de comunidades virtuales, según la tipología:

- Horizontales: promueven la interrelación en general (**Facebook**, **Google+**).
- Verticales: dirigidas a un público específico (**LinkedIn**), promueven una actividad en particular (**Youtube**, **Twitter**).



Figura 2.7: Redes sociales más utilizadas en España. Fuente: iab.

En esta sección, se comentarán las posibilidades publicitarias que ofrecen las principales plataformas sociales: Facebook, Twitter y LinkedIn. Google+ y Youtube serán comentados más adelante.

2.1.3.1 Marketing en Facebook

Facebook Ads ofrece la posibilidad de crear campañas de publicidad segmentadas, de acuerdo con los intereses de los usuarios tanto dentro de la red social como fuera de ella. Permite la promoción de sitios web, páginas de Facebook, eventos o aplicaciones. Estos anuncios pueden contener textos, gráficos o vídeos. Y serán mostrados en el inicio, en el perfil o en las fotos de los usuarios.

Las principales ventajas son: grandes niveles de segmentación, posible viralización del contenido, precios económicos -según el número de clics recibidos-, interacción con los usuarios y medición de los resultados -para optimizar las campañas-.

Las campañas se pueden enfocar según distintos objetivos: atraer personas al sitio web, aumentar las conversiones en el sitio web, promocionar publicaciones, promocionar páginas, aumentar las instalaciones de la aplicación, incrementar la interacción en la aplicación, aumentar los asistentes al evento, conseguir que se soliciten ofertas o aumentar las reproducciones de vídeo.

Facebook facilita la herramienta Power Editor para optimizar las campañas según el objetivo, el tipo de dispositivo o el segmento de mercado al que dirigir las campañas.

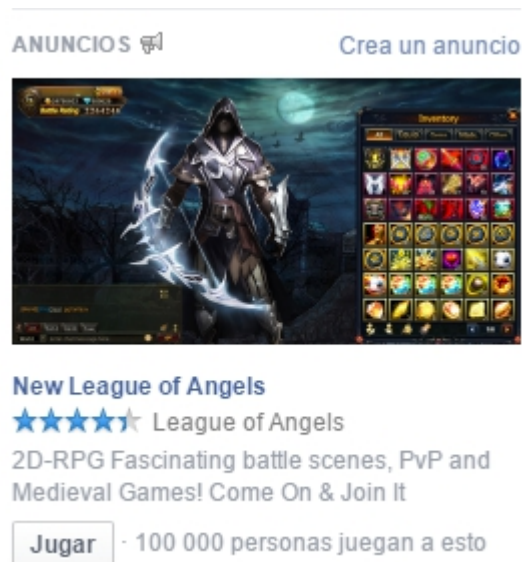


Figura 2.8: Anuncio en Facebook.

2.1.3.2 Marketing en Twitter

Twitter Ads brinda la posibilidad de crear campañas de anuncios para promocionar tweets, cuentas, tendencias o TweetCards; dirigidas a un público determinado según: los intereses, la geografía, el género, el dispositivo, los usuarios similares, la actividad en la red social y fuera de ella. El mensaje cobra mayor relevancia al segmentar los tweets por palabra clave.

Los tweets promocionados son comprados por los anunciantes para dar mayor cobertura a un tweet. Principalmente, para promocionar una marca o incentivar una determinada acción en Twitter. Se pueden encontrar en la parte superior de las búsquedas, en las cronologías de inicio o en los clientes de Twitter. Se factura en función de las acciones que los usuarios hagan en el tweet.



Figura 2.9: Tweet promocionado.

Las cuentas promocionadas pretenden: generar compras, clientes potenciales, descargas y registros, a su vez, incrementar el tráfico web y aumentar el reconocimiento de la marca. Se factura según el número de seguidores (CPF).

Twitter pone a disposición de los anunciantes numerosas herramientas de seguimiento para optimizar las campañas.

2.1.3.3 Marketing en LinkedIn

LinkedIn Ads permite crear campañas de anuncios para publicitarlas en páginas destacadas de la red social. El público -al que van dirigidos los anuncios- puede segmentarse por: cargo, función, sector, ubicación, edad, sexo, nombre y tamaño de la empresa o grupo al que pertenece en LinkedIn.

Su objetivo es incrementar las visitas en un determinado sitio web. Se paga únicamente por los clics recibidos (PPC). Los anuncios siguen el formato de cualquier anuncio de texto y pueden ser mostrados en la página del perfil, en la página de inicio, en el buzón, en la página de resultados de una búsqueda o en la página de un grupo.

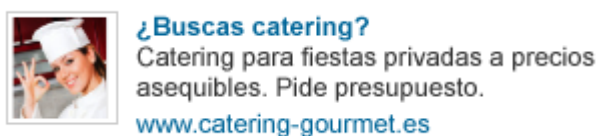


Figura 2.10: Anuncio en LinkedIn.

2.2 Marco Regulador: Leyes de Protección de datos

Tras analizar las distintas opciones existentes para realizar marketing en Internet -cuya tendencia principal es centrarse en la segmentación del mercado según los tipos de usuarios y sus preferencias- parece obvio que es necesario poner un límite a estas prácticas en favor de la privacidad de los individuos.

En este apartado, se describirán las principales leyes de protección de datos españolas y europeas que actualmente se encuentran en vigor, así como el trato que debe darse a los datos de las categorías sensibles.

2.2.1 Leyes Españolas

La privacidad en España está regulada por la Ley Orgánica 15/1999, de 13 de Diciembre, de Protección de Datos de Carácter Personal (**LOPD**), cuya finalidad es garantizar y proteger las libertades públicas y los derechos fundamentales -con respecto al tratamiento de los datos personales de los individuos-, especialmente su honor, intimidad, privacidad personal y familiar. Se pretende regular el tratamiento de los datos y de los ficheros de carácter personal, independientemente del soporte en el que se encuentren, informático o no.

Además deberán tratarse con especial cuidado los datos de **categorías sensibles**, los cuales requieren una mayor privacidad al estar ligados con la intimidad de las personas, sus derechos fundamentales y libertades públicas. En España, su tratamiento se encuentra regulado en el artículo 7 de la LOPD.

En esta categoría se encuentran los datos que revelen la ideología, la afiliación sindical, la religión y las creencias. Así como los datos que hagan referencia al origen racial, a la salud y a la vida sexual. También los datos relativos a las infracciones penales o administrativas.

La Agencia Española de Protección de Datos (AEPD) es la encargada de velar por el cumplimiento de la ley en el ámbito estatal. En caso de incumplir dicha normativa, existen sanciones desde los 900€ -sanciones leves- hasta 600.000€ -sanciones muy graves-. Según la AEPD, el sector con mayor número de sanciones es el de las Telecomunicaciones.

Tras el nacimiento de la web 2.0 y las redes sociales, la AEPD hace una serie de recomendaciones a los proveedores de estos servicios:

- Informar del trato que se va a dar a los datos.
- Establecer políticas de privacidad claras y con perfiles configurables por el usuario.
- Advertir sobre el riesgo de atentado a la intimidad al transferir datos a este medio.
- Informar sobre la compartición de datos de terceros sin su consentimiento.
- Permitir realizar reclamaciones que cubran las cuestiones de protección de datos.
- Establecer plazos máximos para la conservación de datos de cuentas inactivas.
- Limitar los riesgos, en la medida de lo posible.

Por otro lado, la publicidad en medios electrónicos se encuentra regulada específicamente por la Ley 34/2002, de 11 de julio, de servicios de la sociedad de la información y de comercio electrónico (**LSSI**) y por la Ley 32/2003 de noviembre, **General de Telecomunicaciones**.

2.2.2 Leyes Europeas

La **propuesta de reglamento europeo sobre la protección de datos personales** -General Data Protection Regulation (GDPR)- fue aprobada a principios de 2012 por la Comisión, el Parlamento y el Consejo de la Unión Europea. Dichas reformas remplazarán la normativa existente en España, en cuanto a protección de datos -la Ley Orgánica 15/1999, de 13 de Diciembre, de Protección de Datos de Carácter Personal-.

El objetivo de dicha propuesta es revisar las leyes existentes debido al rápido desarrollo tecnológico y al aumento significativo del intercambio de datos. También para unificar criterios sobre el tema en todos los países miembros.

Se establece una política de protección de datos mucho más estricta que las anteriores, en todos los mercados de la Unión Europea. Además se endurecen las sanciones considerablemente.

Se amplía el concepto de datos personales, incluyendo en estos: identificadores únicos como la dirección IP o las cookies. Se incluye la obligación de pedir consentimiento a los usuarios para el tratamiento de sus datos personales, así como avisar del uso de cookies en los distintos sitios web.

En cuanto a la creación de perfiles -cuyo fin sea la publicidad segmentada-, se exige que exista un contrato entre las partes o que el usuario haya dado su consentimiento expreso para poder realizar dicha práctica.

Se espera que para el año **2016** entren en vigor los cambios propuestos. Por otra parte, muchas empresas están en desacuerdo con dichas propuestas. Y piden a la Unión Europea que a cambio incentive a las empresas de alguna forma, ya que la publicidad digital es la base de muchos modelos de negocio en la actualidad.

2.3 Especificación del Problema: Google

Google Inc. es una empresa multinacional estadounidense especializada en productos y servicios relacionados con Internet, software, dispositivos electrónicos, entre otros. Su producto estrella es el **motor de búsqueda** con nombre homónimo.

También ofrece otros servicios como: correo electrónico (**Gmail**), mapas (**Google Maps**, Google Earth), vídeos (**Youtube**), red social (Google+), navegador web (Google Chrome), blogs (Blogger), almacenamiento en la nube (Google Drive), mensajería instantánea (Hangouts), tienda electrónica para dispositivos android (Google Play), publicidad (Google Adwords, Google Adsense), entre muchos otros.

Como ya se ha comentado anteriormente, la cobertura de Google llega al 90-95% de los internautas a nivel mundial. Y la página de su buscador es el sitio web más visitado

de todo el mundo, según el ranking web internacional.

Además, -como se pudo comprobar en el apartado 2.1 Estado del Arte- Google es una de las empresas punteras, en cuanto a **marketing en Internet**. Principalmente, por su amplia cobertura a nivel mundial, situándola en una posición totalmente privilegiada frente a sus competidores.

En primer lugar, las primeras posiciones en los resultados de su motor de búsqueda son el lugar más deseado por cualquiera que pretenda hacer posicionamiento natural de un sitio web (**SEO**). El propio Google ofrece las herramientas de Google Analytics y de webmaster para facilitar el trabajo a los SEOs.

Google también es un lugar estratégico para aquellos que deseen pagar por posicionar sus campañas de anuncios (**SEM**) en uno de los cientos de espacios que ofrece la compañía: en cualquiera de sus servicios o en su extensa red de colaboradores. Entre sus pocos productos de pago se encuentran **Google Adwords** y **Google Adsense** para crear campañas publicitarias dirigidas a un público determinado.

Además, esas campañas se muestran en redes sociales como Youtube o Google+, por lo que Google también cubriría el área de **marketing en redes sociales**.

Ahora bien, Google necesita segmentar las campañas y dirigirlas a un público determinado, ¿cómo se realiza la segmentación de intereses de los usuarios?

La gran mayoría de los productos ofrecidos por Google son gratuitos. Sin embargo, no es del todo cierto, los usuarios de dichos servicios debemos "pagar" un precio, y ese precio es parte de nuestra **privacidad** en la red, la cual es usada para situarnos en un determinado segmento de mercado al que dirigir las rentables campañas publicitarias -citadas anteriormente-.

En ese sentido, la **política de privacidad** de Google no deja lugar a dudas, especifica con total claridad que -prácticamente- todo lo que hagamos con sus servicios es del dominio de Google, y a su vez, fuente para su red de marketing. Y la venta de nuestra privacidad es totalmente consentida por nosotros al usar dichos servicios.

Y si pensamos en los millones de personas que usan al día cualquier servicio de Google, las cifras de recogida de información personal -a nivel mundial- pueden ser simplemente escalofrantes. Y lo más preocupante, legales.

De cualquier forma, Google -en su política de privacidad- garantiza respetar los datos relativos a las **categorías sensibles** -aquellos relacionados con temas como la raza, la religión, la orientación sexual o la salud-. Y aseguran imponer una política similar a sus anunciantes.

En principio, su política de privacidad -así como el trato relativo a los datos de categorías sensibles- cumple con la **normativa vigente** -expuesta en el apartado 2.2

Marco Regulador-. También en cuanto a la exposición de publicidad segmentada según los perfiles, ya que existe un contrato vinculante entre las partes, que garantiza el consentimiento del usuario para tal fin.

En este trabajo, vamos a adentrarnos en las profundidades de Google, con la creación de una herramienta que simule el comportamiento de un usuario al utilizar los servicios básicos de Google. La finalidad principal consiste en el análisis de por qué, cuándo y de qué manera un usuario es introducido en un segmento de mercado determinado -según las acciones que realiza en los distintos servicios-.

El segundo fin es la comprobación de que realmente Google garantiza la privacidad del usuario respecto a los datos de las categorías sensibles. Ya que en más de una ocasión, la multinacional se ha visto en medio de un escándalo por no respetar el trato de estas categorías sensibles. Algunos ejemplos son los siguientes:

- "Google pide perdón por robar datos sensibles de redes Wifi en Alemania". 15/5/2010. Fuente: www.elmundo.es
- "Primera sentencia en España contra Google por protección de datos sensibles". 16/10/2014. Fuente: www.notariallopis.es

En los siguientes apartados, se introducirán los servicios que serán estudiados: Buscador, Gmail, Google Maps y Youtube. También los productos publicitarios: Google Adwords y Google Adsense. Finalmente, se introducirá la configuración de anuncios de Google.

2.3.1 Servicios de Google

2.3.1.1 Buscador

Es el motor de búsqueda web más utilizado del mundo, basado en el algoritmo exclusivo PageRank. Dicho buscador recibe millones de consultas diarias a través de sus diferentes servicios.

Entre los cuales, se incluyen: sinónimos, zonas horarias, previsiones meteorológicas, cotizaciones de bolsa, cartelera cinematográfica, resultados deportivos, etc... Además, permite realizar traducciones, cálculos, conversiones de unidades y de monedas, entre muchas otras funcionalidades extras. Google también proporciona muchas opciones de búsqueda avanzada para refinar las consultas y hacerlas totalmente personalizadas.

Como producto estrella ocupa un lugar destacado en las estrategias de marketing online de la multinacional. En la figura 2.5, se pueden apreciar los lugares designados por Google para el posicionamiento SEM -para aquellos anuncios contratados en las campañas de Google Adwords-.



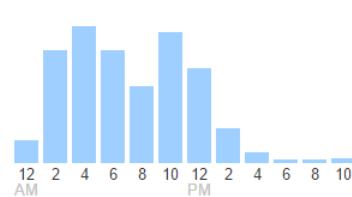
Figura 2.11: Página principal de Google, cuyo Doodle (logotipo) es cambiante, según diversos eventos relevantes.

Por otro lado, en el **historial de búsquedas** de nuestra cuenta, se puede comprobar que Google registra todas las consultas que hacemos: detallando la fecha, así como los anuncios y las páginas en las que hemos hecho clic tras utilizar el servicio. Para después, generar gráficas con nuestras **tendencias** de búsqueda.



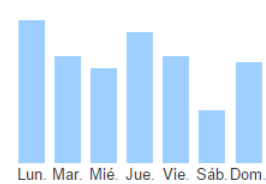
Figura 2.12: Resumen del historial de búsquedas de Google.

Actividad de búsqueda por hora

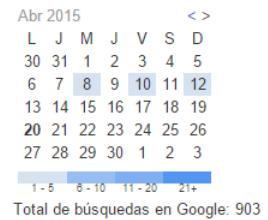


<< Mostrar menos tendencias

Actividad de búsqueda por día



Actividad



Mayor número de clics

1. eBiblio Madrid
2. YouTube
3. <http://es.lightspeedpanel.com/>
4. <http://www.hotmail.com/?mkt=es-es>
5. <http://translate.google.com/?hl=es>

Consultas principales

1. hotmail
2. ebiblio
3. youtube
4. trad
5. gmail

Sitios principales

1. www.youtube.com
2. es.wikipedia.org
3. www.hotmail.com
4. madrid.ebiblio.es
5. es.lightspeedpanel.com

Actividad de búsqueda por mes

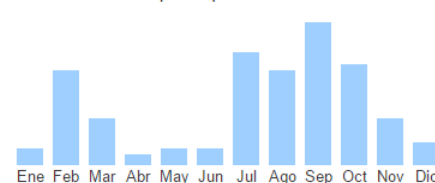


Figura 2.13: Tendencias del historial de búsquedas de Google.

2.3.1.2 Gmail

Servicio de correo electrónico famoso por sus innovaciones tecnológicas y su capacidad. En Noviembre de 2012, consiguió rebasar a Outlook en cuanto a número de usuarios en todo el mundo. Además, se encuentra en el punto de mira por supuestas violaciones de la privacidad de los usuarios, al colaborar con los servicios de inteligencia de Estados Unidos.

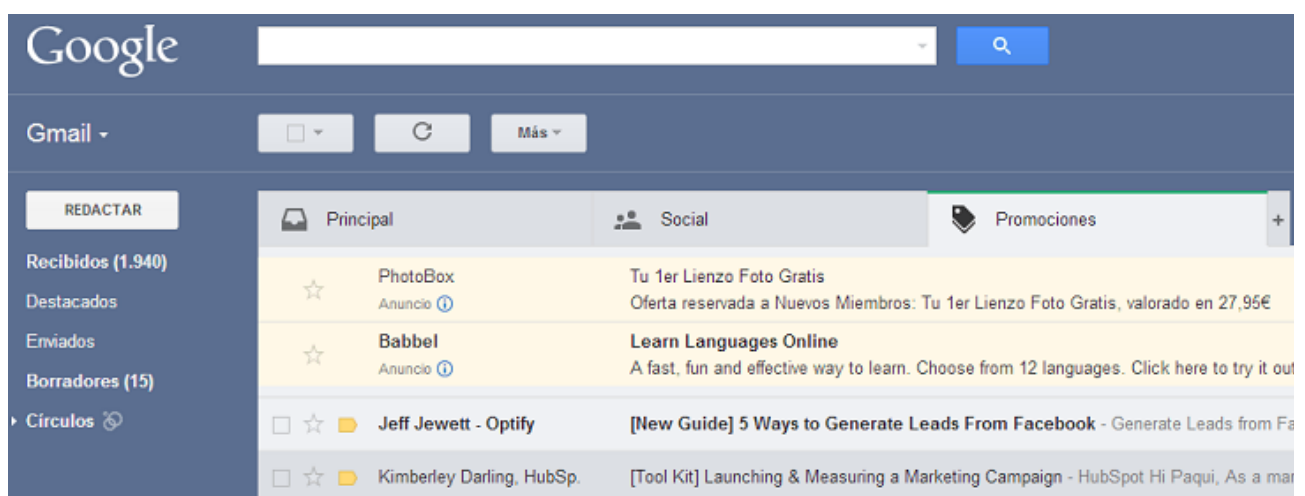


Figura 2.14: Anuncios de Google Adwords en Gmail.

En la **política de privacidad**, se nos informa que los correos electrónicos intercambiados en Gmail son parte del entramado para crear los perfiles basados en intereses. Aunque esta segmentación sea realizada por bots y se asegura que nadie lee nuestra correspondencia, de cualquier manera, es clasificada en el **historial** de nuestra cuenta.



Figura 2.15: Resumen del historial de Gmail.

2.3.1.3 Google Maps

Servicio de mapas en la web: ofrece mapas, fotografías tomadas por satélite, rutas entre distintas ubicaciones e incluso imágenes a pie de calle -gracias a Google Street View-. También se encuentra en el punto de mira por estar -supuestamente- ayudando a desarrollar sistemas de vigilancia masivos con el uso de los smartphones.

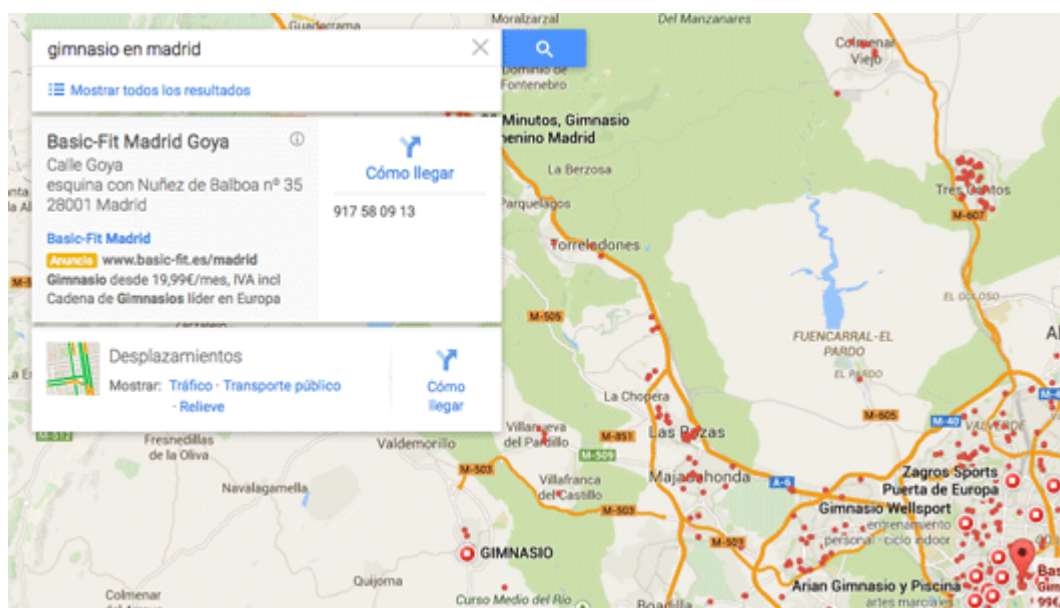


Figura 2.16: Anuncios de Google Adwords en Google Maps.

Comparte el **historial de búsquedas** con otros servicios como el buscador, el buscador de vídeos, de imágenes, etc...

2.3.1.4 Youtube

Sitio web de alojamiento de vídeos, alberga un reproductor de vídeo basado en Adobe Flash o en el estándar HTML5. Aloja fragmentos de películas, programas de televisión y vídeos musicales, entre otros. Es el sitio de vídeos más popular de la red.

Además de anuncios de texto, permite la posibilidad de mostrar anuncios de vídeo en las campañas de Google Adwords para un público determinado (Figura 2.4). También facilita la herramienta Analytics de YouTube para gestionar los resultados de dichas campañas.

Existe un **historial de búsquedas** y otro **historial de reproducciones**, en los que se almacenan los datos relativos al uso del servicio.

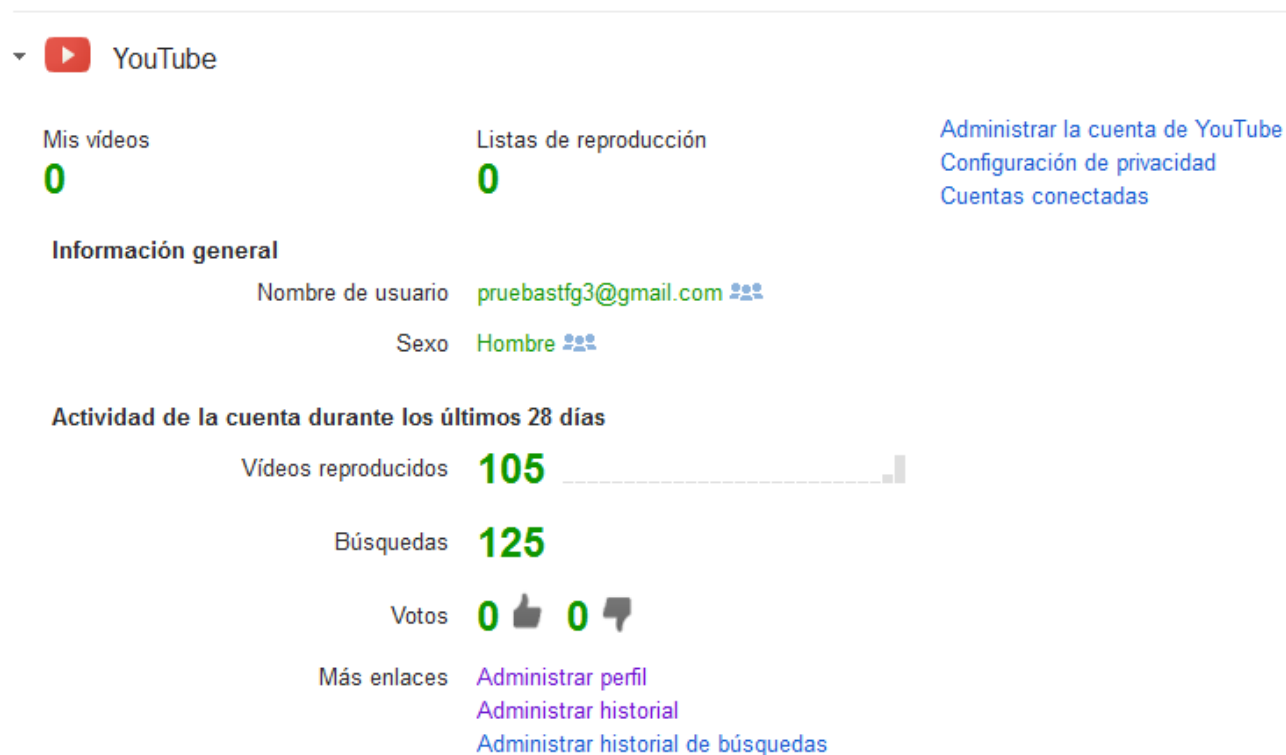


Figura 2.17: Resumen del historial de búsquedas y del historial de reproducciones en Youtube.

2.3.2 Servicios de publicidad de Google

2.3.2.1 Google Adwords

Servicio publicitario que ofrece Google a los anunciantes para crear campañas de anuncios dirigidos a un segmento determinado de mercado (posicionamiento SEM). Es la principal fuente de ingresos de la compañía.

Estos anuncios pueden aparecer en los sitios pertenecientes a su red de búsqueda -servicios pertenecientes a Google, descritos en el apartado 2.3.1 Servicios de Google-, y también en los sitios web asociados a su red de display. El formato de los anuncios en la red de display es mucho más flexible -vídeo, imágenes, banners dinámicos, etc...- que en la red de búsqueda -texto-.

El pago de los anuncios se realiza según el número de clics realizados, siguiendo el formato de pago por clics (PPC). El precio de cada clic es variable, se subasta en función de la oferta y de la demanda, así como de otros factores relativos a la calidad y la relevancia del mismo.

Cada campaña puede tener uno o varios grupos de anuncios. A su vez, cada grupo de anuncios puede tener uno o varios anuncios, palabras clave, ubicaciones y otros datos para realizar la segmentación del mercado. La herramienta Google Analytics también es válida para medir el rendimiento de dichas campañas.

Resulta curioso que algunas **palabras clave**, así como algunas **categorías** del árbol de intereses -que Google ofrece a los anunciantes- se encuentren dentro de las **categorías sensibles**. (Para más información: <https://support.google.com/adwords/answer/156178?hl=es>). Más adelante profundizaremos en este tema.

2.3.2.2 Google Adsense

Es el otro servicio de marketing de Google para la creación de campañas publicitarias personalizadas en sitios web de particulares. Estos anuncios se muestran en páginas web en forma de banners, imágenes, vídeos o texto -dentro de un marco HTML que hace referencia a Google-.

Los ingresos de esta publicidad son compartidos entre Google y los administradores de las páginas web en las que aparecen dichos anuncios. El pago se realiza por cada clic generado desde los anuncios del sitio web. Además el fraude es perseguido rigurosamente por Google, para aquellos que quieran lucrarse indiscriminadamente, especialmente a través de un PPF (Puntos Publicitarios Fraudulentos).

Todos los ingresos de Google Adsense provienen de los anunciantes de Google Adwords.



Figura 2.18: Anuncios de Google AdSense.

2.3.3 Configuración de Anuncios de Google

Una vez comentados los tipos de anuncios que nos podemos encontrar -a la hora de utilizar los servicios de Google-, así como los tipos de campañas personalizadas que éste ofrece a sus anunciantes. Solo queda comentar cómo se realiza la segmentación de anuncios por intereses.

Los factores influyentes para la personalización de los anuncios son: los datos demográficos, la búsqueda actual, el historial web, los sitios web visitados que se anuncian a través de Google, las interacciones con los servicios de Google y con los anuncios, los mensajes de Gmail en la bandeja de entrada y de salida, así como los ajustes de la configuración de anuncios.

Google permite al usuario inhabilitar los anuncios basados en intereses o bloquear a anunciantes concretos pero eso no exime al usuario de seguir viendo publicidad mientras utilice sus servicios.

En la página de configuración de anuncios de nuestra cuenta de Google (Figura 2.19), podemos ver el tipo de perfil que Google nos crea mientras navegamos. También podemos modificar la información que ahí aparezca, ya sea añadiendo o eliminando intereses.

En dicha página se pueden apreciar dos secciones:

- **Anuncios en Google:** Aparecen los datos relativos a nuestra cuenta de Google (sexo y edad), así como los intereses en función de nuestra actividad anterior en Google.
- **Anuncios Google en la Web:** Aparecen el sexo, la edad y los intereses en función de las páginas web que hayamos visitado.

Todas las **categorías de intereses** que Google utiliza -para la segmentación de anuncios basados intereses de los usuarios-, se pueden encontrar en la siguiente página web: <https://support.google.com/ads/answer/2842480?hl=es>

Configuración de los anuncios Google

Soluciones publicitarias habilita contenidos y servicios web gratuitos. Esta configuración permite controlar los tipos de anuncios Google que ves.

	Anuncios en Google	Anuncios Google en la Web [?]
	 Búsqueda  Gmail  YouTube  Google Maps	 Anuncios Google en la Web
Sexo	Mujer Visitar tu perfil de Google	N/D
Edad	25-34 Visitar tu perfil de Google	N/D
Idiomas	N/D	N/D
Intereses	Desconocido Editar De tu actividad anterior en Google	N/D

Figura 2.19: Configuración de anuncios de Google.

3. DISEÑO DE LA SOLUCIÓN TÉCNICA

3.1 Análisis de Requisitos

En este apartado, se detallarán los requisitos que debe cumplir la solución técnica a desarrollar. Cada requisito se mostrará en una tabla, indicando el tipo, una descripción del mismo, así como el nivel (alto, medio, bajo) de prioridad, necesidad y estabilidad de cada uno.

RQ-001	Automatización del Buscador	
	Tipo	Funcional
	Descripción	El sistema debe permitir realizar la búsqueda de una o varias palabras en Google, así como la interacción con los resultados mostrados -de manera automática-.
	Prioridad	Alta
	Necesidad	Alta
	Estabilidad	Alta

Tabla 3.1: RQ-001: Automatización del Buscador.

RQ-002	Automatización de Gmail	
	Tipo	Funcional
	Descripción	El sistema debe permitir el envío de correos electrónicos -de manera automática-.
	Prioridad	Alta
	Necesidad	Alta
	Estabilidad	Alta

Tabla 3.2: RQ-002: Automatización de Gmail.

RQ-003	Automatización de Google Maps	
	Tipo	Funcional
	Descripción	El sistema debe permitir realizar la búsqueda de una o varias palabras en Google Maps, así como la interacción con los resultados mostrados -de manera automática-.
	Prioridad	Alta
	Necesidad	Alta
	Estabilidad	Alta

Tabla 3.3: RQ-003: Automatización de Google Maps.

RQ-004	Automatización de Youtube	
	Tipo	Funcional
	Descripción	El sistema debe permitir realizar la búsqueda de una o varias palabras en Youtube, así como la interacción con los resultados mostrados -de manera automática-.
	Prioridad	Alta

Necesidad	Alta
Estabilidad	Alta

Tabla 3.4: RQ-004: Automatización de Youtube.

RQ-005	Automatización de Login
Tipo	Funcional
Descripción	El sistema debe permitir la realización del login y la comprobación de que el usuario ha podido acceder a su cuenta de Google correctamente -de manera automática-.
Prioridad	Alta
Necesidad	Alta
Estabilidad	Alta

Tabla 3.5: RQ-005: Automatización de Login.

RQ-006	Automatización de Logout
Tipo	Funcional
Descripción	El sistema debe permitir la realización del logout de la cuenta de Google del usuario -de manera automática-.
Prioridad	Media
Necesidad	Media
Estabilidad	Alta

Tabla 3.6: RQ-006: Automatización de Logout.

RQ-007	Extracción Inteligente
Tipo	Funcional
Descripción	El sistema debe permitir la extracción de información relevante de la configuración de anuncios de la cuenta de Google del usuario -de manera automática-.
Prioridad	Alta
Necesidad	Alta
Estabilidad	Alta

Tabla 3.7: RQ-007: Extracción Inteligente.

RQ-008	Almacenamiento General
Tipo	Funcional
Descripción	El sistema debe permitir el almacenamiento de información relevante a la prueba en una base de datos local.
Prioridad	Alta
Necesidad	Alta
Estabilidad	Alta

Tabla 3.8: RQ-008: Almacenamiento General.

RQ-009	Almacenamiento Individual
Tipo	Funcional
Descripción	El sistema debe permitir el almacenamiento de información extraída en una base de datos local.
Prioridad	Alta
Necesidad	Alta

Estabilidad	Alta
--------------------	------

Tabla 3.9: RQ-009: Almacenamiento Individual.

RQ-010	Personalización de la Prueba
Tipo	Funcional
Descripción	El sistema debe permitir al usuario personalizar la prueba, de manera que se ajuste lo máximo posible a sus necesidades, en cuanto a: duración, número de iteraciones, tipo de servicio a utilizar, cuenta origen, cuenta destino, envío o recepción de correos electrónicos. Así como la procedencia de los datos, las palabras (los asuntos) y los contenidos a utilizar, el tipo de interacción con los servicios y el tipo de rastreo en la navegación.
Prioridad	Media
Necesidad	Alta
Estabilidad	Alta

Tabla 3.10: RQ-010: Personalización de la Prueba.

RQ-011	Selección de Servicio
Tipo	Funcional
Descripción	El sistema debe permitir seleccionar al usuario uno o más servicios.
Prioridad	Alta
Necesidad	Alta
Estabilidad	Alta

Tabla 3.11: RQ-011: Selección de Servicio.

RQ-012	Personalización de Correo Electrónico
Tipo	Funcional
Descripción	El sistema debe permitir al usuario seleccionar la cuenta origen y destino de los correos electrónicos, así como la posibilidad de enviarlos o recibirlos.
Prioridad	Media
Necesidad	Alta
Estabilidad	Alta

Tabla 3.12: RQ-012: Personalización de Correo Electrónico.

RQ-013	Selección de Categorías
Tipo	Funcional
Descripción	El sistema debe permitir al usuario seleccionar un nivel, o un rango de niveles, de las categorías de intereses proporcionadas por Google.
Prioridad	Media
Necesidad	Alta
Estabilidad	Alta

Tabla 3.13: RQ-013: Selección de Categorías.

RQ-014	Gestión de Ficheros
Tipo	Funcional
Descripción	El sistema debe permitir la selección, la lectura y la extracción de información relevante de un fichero de texto seleccionado por el usuario.
Prioridad	Media
Necesidad	Alta
Estabilidad	Alta

Tabla 3.14: RQ-014: Gestión de Ficheros.

RQ-015	Selección de Palabra
Tipo	Funcional
Descripción	El sistema debe permitir seleccionar al usuario una o más palabras (asuntos), procedentes de las categorías de intereses o de un fichero de texto.
Prioridad	Alta
Necesidad	Alta
Estabilidad	Alta

Tabla 3.15: RQ-015: Selección de Palabra.

RQ-016	Selección de Contenido
Tipo	Funcional
Descripción	El sistema debe permitir seleccionar al usuario uno o más contenidos, procedentes de las categorías de intereses o de un fichero de texto.
Prioridad	Alta
Necesidad	Alta
Estabilidad	Alta

Tabla 3.16: RQ-016: Selección de Contenido.

RQ-017	Interacción con los Servicios
Tipo	Funcional
Descripción	El sistema debe permitir seleccionar al usuario si quiere profundizar en la prueba, interactuando con los resultados de los servicios.
Prioridad	Media
Necesidad	Alta
Estabilidad	Alta

Tabla 3.17: RQ-017: Interacción con los Servicios.

RQ-018	Rastreo
Tipo	Funcional
Descripción	El sistema debe permitir seleccionar al usuario si quiere ser rastreado mientras navega.
Prioridad	Media
Necesidad	Media
Estabilidad	Alta

Tabla 3.18: RQ-018: Rastreo.

RQ-019 Tipo	Validación de Datos Introducidos
	Funcional
Descripción	El sistema debe permitir la validación de los datos introducidos por el usuario, a la hora de personalizar la prueba (duración, número de iteraciones, cuenta origen, cuenta destino), incluido el fichero de texto seleccionado.
Prioridad	Media
Necesidad	Alta
Estabilidad	Alta

Tabla 3.19: RQ-019: Validación de Datos Introducidos.

RQ-020 Tipo	Sucesión de Pruebas
	Funcional
Descripción	El sistema debe permitir realizar otra nueva prueba tras la conclusión de la anterior.
Prioridad	Media
Necesidad	Media
Estabilidad	Alta

Tabla 3.20: RQ-020: Sucesión de Pruebas .

RQ-021 Tipo	Gestión de Errores
	Funcional
Descripción	El sistema debe tratar e informar al usuario sobre cualquier incidencia acontecida en el transcurso de la prueba.
Prioridad	Media
Necesidad	Alta
Estabilidad	Alta

Tabla 3.21: RQ-021: Gestión de Errores.

RQ-022 Tipo	Interfaz de Usuario Intuitiva
	No Funcional
Descripción	El sistema debe proporcionar una interfaz de usuario sencilla e intuitiva para facilitar la utilización a los usuarios.
Prioridad	Media
Necesidad	Media
Estabilidad	Alta

Tabla 3.22: RQ-022: Interfaz de Usuario Intuitiva.

RQ-023 Tipo	Selección de Opciones en la Intefaz de Usuario
	No Funcional
Descripción	El sistema debe proporcionar una interfaz de usuario que facilite la sencilla personalización de las pruebas por parte del usuario.
Prioridad	Media
Necesidad	Media
Estabilidad	Alta

Tabla 3.23: RQ-023: Selección de Opciones en la Interfaz de Usuario.

RQ-024 Tipo Descripción Prioridad Necesidad Estabilidad	Valores por Defecto en la Intefaz de Usuario
	No Funcional
	El sistema debe proporcionar una serie de valores por defecto para orientar al usuario en la personalización de las pruebas.
	Media
	Media

Tabla 3.:24 RQ-024: Valores por Defecto en la Interfaz de Usuario.

RQ-025 Tipo Descripción Prioridad Necesidad Estabilidad	Botón de Información en la Intefaz de Usuario
	No Funcional
	El sistema debe proporcionar una guía básica, en la interfaz de usuario, para orientar a éste sobre el uso de la herramienta y la selección de las distintas opciones.
	Media
	Media

Tabla 3.25: RQ-025: Botón de Información en la Interfaz de Usuario.

RQ-026 Tipo Descripción Prioridad Necesidad Estabilidad	Velocidad del Sistema
	No Funcional
	El sistema debe optimizar las búsquedas (o el envío de correos electrónicos), así como la extracción y el almacenamiento de información para agilizar el desarrollo temporal de las pruebas.
	Alta
	Alta

Tabla 3.26: RQ-026: Velocidad del Sistema.

RQ-027 Tipo Descripción Prioridad Necesidad Estabilidad	Rendimiento del Sistema
	No Funcional
	El sistema debe consumir los recursos mínimos necesarios para mayor rendimiento.
	Alta
	Alta

Tabla 3.27: RQ-027: Rendimiento del Sistema.

RQ-028 Tipo Descripción Prioridad Necesidad Estabilidad	Seguridad del Sistema
	No Funcional
	El sistema no debe ser un riesgo para el equipo que lo utilice.
	Alta
	Alta

Tabla 3.28: RQ-028: Seguridad del Sistema.

RQ-029	Portabilidad del Sistema
Tipo	No Funcional
Descripción	El sistema debe poder usarse en cualquier plataforma.
Prioridad	Alta
Necesidad	Alta
Estabilidad	Alta

Tabla 3.29: RQ-029: Portabilidad del Sistema.

RQ-030	Escalabilidad del Sistema
Tipo	No Funcional
Descripción	El sistema debe poder añadir nuevas funcionalidades.
Prioridad	Media
Necesidad	Alta
Estabilidad	Alta

Tabla 3.30: RQ-030: Escalabilidad del Sistema.

3.2 Diseño de la Solución Técnica

En este apartado, se comentarán los aspectos relativos al diseño de la solución técnica: la arquitectura elegida, así como las herramientas y tecnologías a utilizar en el desarrollo de la misma.

3.2.1 Arquitectura

El patrón de arquitectura software elegido -para la solución- es el **Modelo-Vista-Controlador** (MVC). Este tipo de arquitectura separa los datos (modelo), la lógica de negocio (controlador) y la interfaz de usuario (vista) en distintos módulos, facilitando la separación de conceptos, la reutilización de código, así como la escalabilidad del sistema para futuras mejoras y ampliaciones. A continuación, se presentan cada uno de los módulos de la solución elegida:

- **Vista:** Presenta el modelo y la lógica de negocio en un formato adecuado para que el usuario pueda interactuar con el sistema. En nuestro caso, consistirá en la interfaz gráfica de usuario (GUI) de una aplicación de escritorio. En futuras ampliaciones, podría adaptarse dicha interfaz gráfica de usuario como página web o aplicación móvil.

Por un lado, el usuario podrá comunicarse con el sistema -personalizando las pruebas según sus necesidades-. Por otro lado, el sistema se comunicará con el usuario -informando del estado de la prueba o de cualquier incidencia que ocurra en el transcurso de la misma-. También el sistema guiará al usuario en la personalización de dichas pruebas con una serie de valores por defecto y una sección de información sobre el funcionamiento de la herramienta.

La tecnología que se empleará en dicha interfaz gráfica de usuario será **java**, más concretamente, la biblioteca gráfica **Swing**. El empleo de java aportará portabilidad al

sistema, pudiendo utilizarse sobre cualquier plataforma.

La elección de java se debe a las posibilidades que ofrece, así como la compatibilidad con Selenium. Otros lenguajes que podrían haberse utilizado por ser compatibles con Selenium son: C#, Ruby, Groovy, Perl, Php y Python. Además la interfaz gráfica de usuario podría desarrollarse con JavaFX pero se ha optado por Swing por tener amplia experiencia en el desarrollo con dicha biblioteca.

- **Controlador:** Responde peticiones provenientes de la vista y hace peticiones al modelo para la gestión de datos. También puede enviar información a la vista procedente del modelo. Es el intermediario entre la vista y el modelo.

En nuestro caso, será el encargado de simular el comportamiento de un usuario en los distintos servicios de Google. Sus funciones principales serán: gestionar las pruebas pedidas por el usuario en la vista de manera automática, extraer la información relevante sobre la configuración de anuncios de manera inteligente, gestionar las peticiones de almacenamiento para que el modelo guarde la información general de la prueba, así como la información relevante extraída en la base de datos local. Y también gestionar las peticiones de consulta de información de las categorías de intereses -proporcionadas por Google-. Finalmente, se informará a la vista sobre el estado de la prueba y las posibles incidencias ocurridas en ésta.

Como ya se comentó en la vista, la principal tecnología empleada será **java**. Para la automatización de las pruebas y la extracción de información relevante, se empleará **Selenium WebDriver**. Y para la localización de elementos HTML dinámicos en las distintas páginas web -pertenecientes a Google-, se empleará **Xpath**. Ambos soportados por java. Finalmente, el navegador web utilizado -para la simulación del comportamiento de usuario- será **Mozilla Firefox**.

Podrían haberse utilizado otros navegadores compatibles con Selenium como: Internet Explorer o Google Chrome. Sin embargo, se ha optado por Mozilla Firefox debido a su compatibilidad con Selenium IDE.

- **Modelo:** Corresponde con la representación de la información del sistema, gestiona todos los accesos y modificaciones a dicha información, así como las peticiones realizadas por la lógica de negocio.

En nuestro caso, corresponderá con la propia base de datos local, así como con la gestión de las peticiones. Se recibirán peticiones de almacenamiento por parte del controlador -tanto para almacenar información general relacionada con la prueba como con información individual extraída de la configuración de anuncios-, así como peticiones de consulta de información de las categorías de intereses -proporcionadas por Google-.

La base de datos estará dividida en 4 secciones: la primera albergará los datos de las categorías de intereses de Google, la cual no será modificada, simplemente consultada por el controlador.

ARQUITECTURA DEL SISTEMA

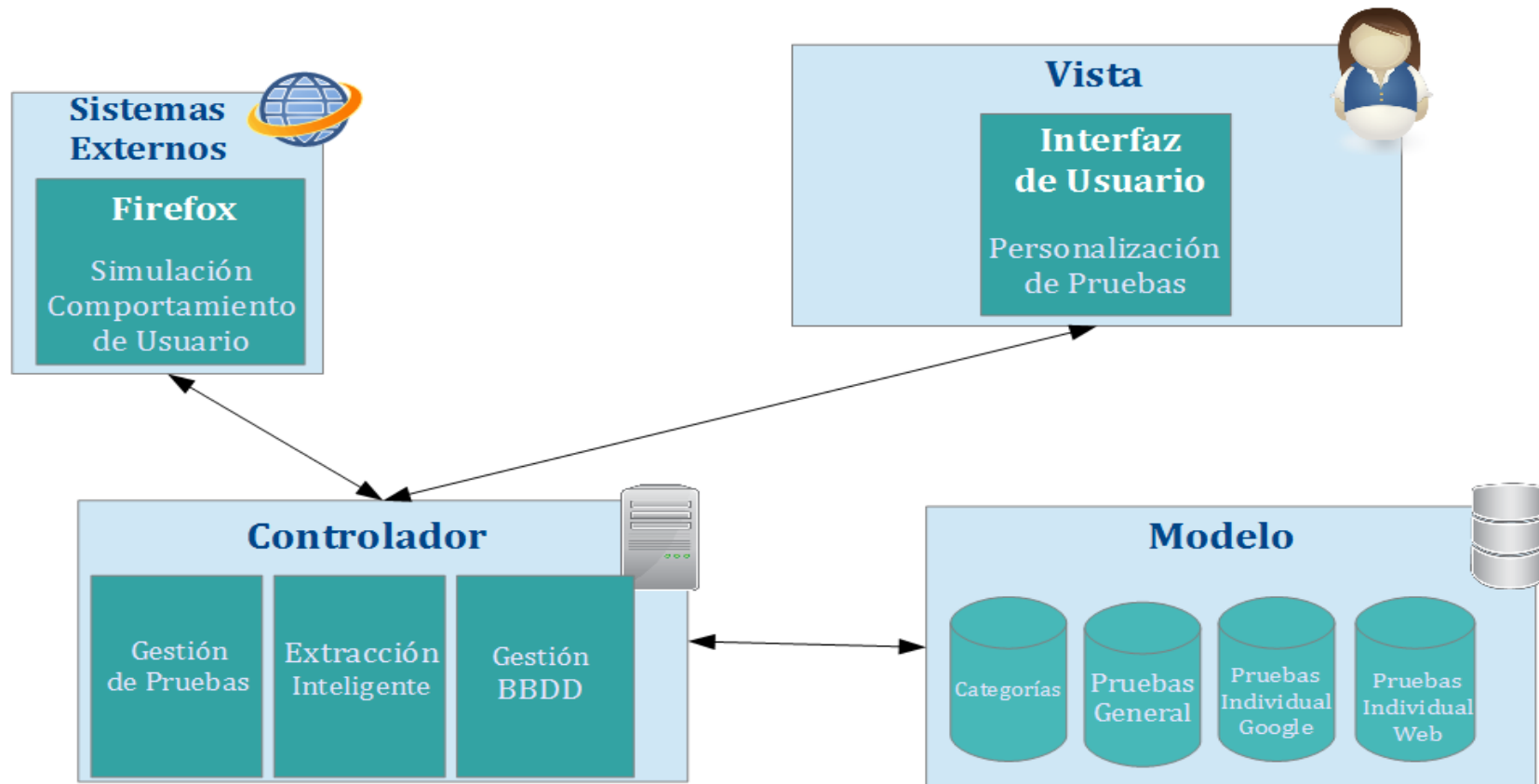


Figura 3.1: Arquitectura del Sistema

La segunda sección corresponderá con los datos relacionados con la información general de las pruebas. La tercera y cuarta corresponderán con la información relevante extraída de la configuración de anuncios de Google, una para los anuncios en Google y otra para los anuncios de Google en la web. Las tres últimas secciones serán modificadas a petición del controlador.

El gestor de base de datos empleado será **MySQL**, el cual aportará concurrencia a la utilización de la misma. Escogido frente a otros por su sencillez, flexibilidad y velocidad, así como su compatibilidad con Java.

En la figura 3.1, se puede ver un diagrama con la arquitectura del sistema que se acaba de describir.

3.2.2 Tecnologías

En este apartado, se van a detallar las principales herramientas y tecnologías que serán utilizadas en la implementación del proyecto. También las características de éstas y su aportación al desarrollo del mismo.

3.2.2.1 Java

Java es un lenguaje de programación de propósito general, concurrente y orientado a objetos, que fue diseñado para tener las mínimas dependencias de implementación. Es uno de los lenguajes más populares en uso, especialmente para las aplicaciones cliente-servidor web.

Sus características principales son:

- Orientado a objetos: Incluye técnicas como encapsulación, herencia, polimorfismo, etc...
- Amplia disponibilidad de bibliotecas: Existe un extenso conjunto de clases a utilizar, con las que se puede realizar prácticamente cualquier tipo de aplicación.
- Distribuido: Facilita la creación de aplicaciones distribuidas.
- Interpretado y compilado: Es compilado porque su código fuente se transforma en código máquina, y es interpretado ya que ese código máquina se puede ejecutar sobre cualquier máquina que tenga el correspondiente interprete.
- Robusto: Permite la creación de aplicaciones altamente fiables.
- Seguro: Garantiza una serie de barreras de seguridad en tiempo real.
- Independiente de la arquitectura: Está diseñado para ejecutarse en distintas

plataformas, gracias al compilador y al interprete de java.

- Portable: La máquina virtual de java (JVM) facilita la portabilidad de las aplicaciones, independientemente de la arquitectura.
- Alto rendimiento.
- Concurrente: Facilita la creación aplicaciones con múltiples hilos.
- Dinámico: La ejecución se produce en tiempo real de manera dinámica.

La elección de java se debió -principalmente- a las posibilidades que ofrece el propio lenguaje en cuanto a **portabilidad**, a **conurrencia**, a **robustez**, entre otras. Así como la **compatibilidad** para usarse con otras bibliotecas, propias de java como Swing o externas como Selenium WebDriver, Xpath y MySQL.

Swing será la biblioteca gráfica a utilizar en el desarrollo de la interfaz gráfica de usuario. Sigue un modelo simple de programación por hilos y cumple con las características propias del lenguaje java. Además, permite la sencilla personalización del diseño gráfico de las distintas ventanas.

En futuras ampliaciones, podría migrarse la interfaz gráfica de usuario a la tecnología JavaFX, la cual permite diseñar ventanas más modernas, así como la posibilidad de añadir interacción multimedia a las mismas. Además permite ejecutar las aplicaciones en distintos soportes como: web, móviles, aplicaciones de escritorio, televisiones, consolas de videojuegos, reproductores Blu-ray, entre otros.

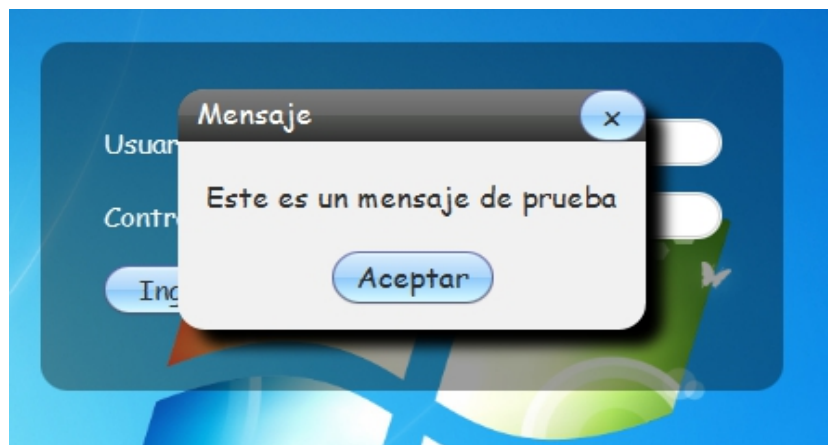


Figura 3.2: Ejemplo de JavaFX. Fuente: tutorialjavaafx.blogspot.com.es

En cuanto al entorno de desarrollo integrado (IDE), el elegido será **Eclipse** por la compatibilidad con java y las posibilidades que ofrece a la hora de añadir distintas bibliotecas. Existe una variedad de plugins muy amplia en Eclipse, dotando al entorno de desarrollo de un gran número de funcionalidades. Incluye: un editor de código fuente, un compilador e interprete, un depurador, entre muchas otras herramientas que facilitan el desarrollo de los proyectos de software. Netbeans también podría haber sido elegido como

entorno de desarrollo. Sin embargo, está más limitado para añadir plugins que Eclipse.

3.2.2.2 Selenium

Selenium consiste en un entorno para automatizar pruebas de aplicaciones web. Dicho entorno proporciona una serie de herramientas para la automatización, entre las cuales: grabar y reproducir pruebas -sin utilizar ningún lenguaje de scripting-, así como exportar las grabaciones en diferentes lenguajes de programación como Java, C#, Ruby, Groovy, Perl, Php y Python. Y también proporciona una serie de librerías para ser utilizadas en los distintos clientes.

Estas pruebas pueden ejecutarse en cualquier navegador moderno, independientemente de la plataforma utilizada. Selenium también permite la posibilidad de comprobar si los resultados son los esperados para una determinada prueba.

Para la localización inteligente de elementos HTML, se utilizan campos como el id o el nombre, así como el lenguaje Xpath. Este lenguaje permite buscar y seleccionar un elemento en una estructura jerárquica de XML. De esta manera, al crear expresiones regulares, podemos encontrar y seleccionar -de manera automática- los distintos elementos de una página web, siempre que estos elementos no cambien su localización.

Las principales herramientas son: Selenium IDE, Selenium Client API, Selenium Remote Control, Selenium WebDriver y Selenium Grid. A continuación, se analizan cada una de ellas:

- **Selenium IDE:** Entorno de desarrollo integrado para la realización de pruebas con Selenium. Consiste en un complemento (extensión) para el navegador web **Mozilla Firefox**.

Permite la grabación, edición, depuración y reproducción de distintas acciones. Estos scripts son generados en el lenguaje propio Selenese, el cual proporciona una serie de comandos para ejecutar acciones sobre el navegador como: hacer clic en un enlace, seleccionar una opción, verificar algún texto, etc...

Posteriormente, estos scripts se pueden exportar a los lenguajes de programación -citados anteriormente- y adaptarse según las necesidades.

En la figura 3.3, puede verse la grabación de distintas acciones en la interfaz de Selenium IDE.

- **Selenium Client API:** Conjunto de librerías que ofrece Selenium -para la realización de pruebas en aplicaciones cliente-, en lenguajes de programación distintos a Selenese. Los clientes se comunican con Selenium mediante la llamada a métodos de Selenium Client API. Actualmente se encuentra disponible en los siguientes lenguajes: **Java**, C#, Ruby y Python.

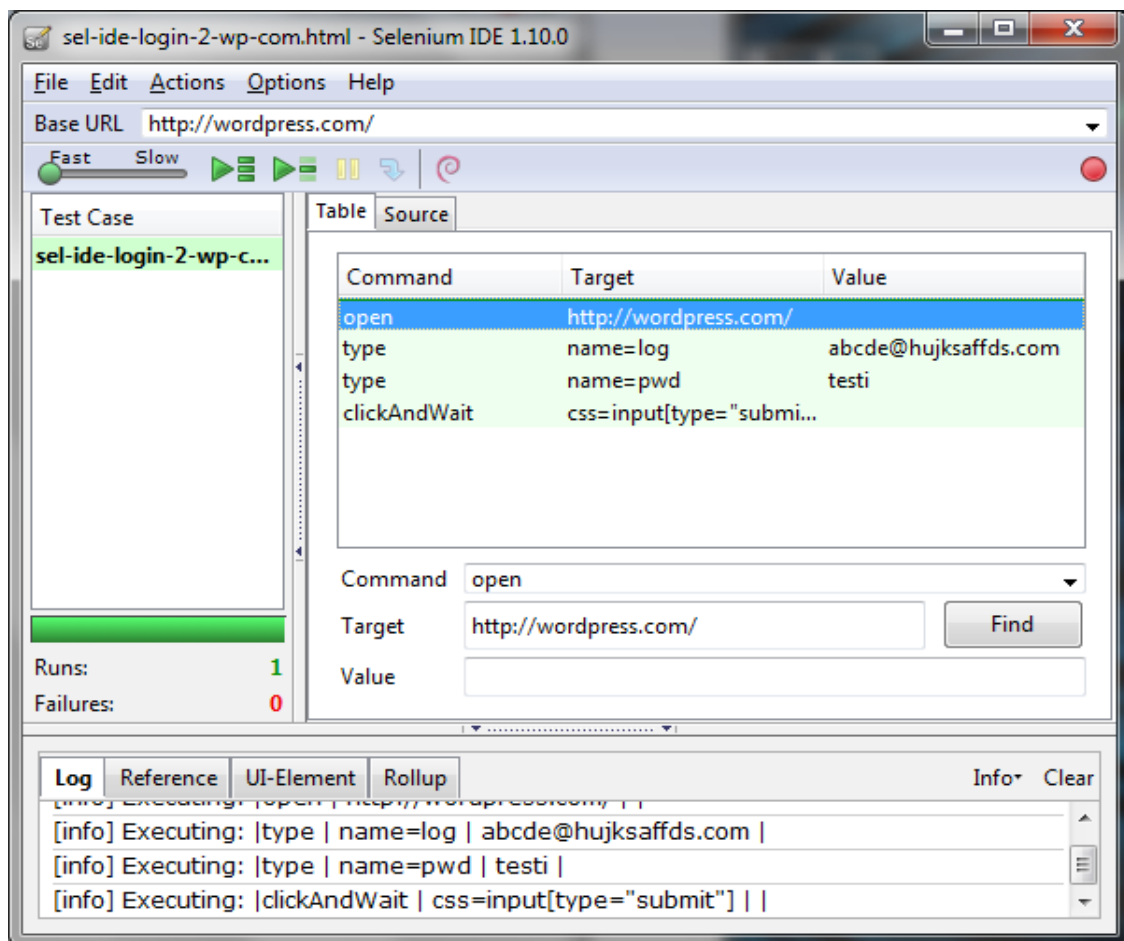


Figura 3.3: Interfaz de Selenium IDE. Fuente: docs.seleniumhq.org

- **Selenium Remote Control (RC):** Servidor escrito en java que interactúa -a base de comandos- con el navegador vía HTTP. Permite la automatización de pruebas -en entornos web- en los distintos lenguajes de programación permitidos. Selenium RC provee una serie de controladores para facilitar la escritura de las pruebas en los diferentes lenguajes. Actualmente se encuentra obsoleto, fue sustituido por Selenium WebDriver.

- **Selenium WebDriver:** Es el sucesor de Selenium RC. Crea controladores específicos de navegador, y dicho controlador envía comandos (seanese o mediante el API de cliente) al navegador y recoge los resultados. La mayoría de navegadores aceptan el uso de Selenium WebDriver aunque existe un controlador para simular un navegador web (HtmlUnit). No se requiere de servidor para ejecutar las pruebas, simplemente se crean instancias de navegador y se controlan.

Selenium 2.0 provee una serie de bloques básicos, permitiendo al desarrollador adaptar las pruebas a su correspondiente lenguaje. Éste pretende convertirse en un estándar de Internet, siguiendo los pasos del proyecto Watir (Web Application Test In Ruby).

- **Selenium Grid:** Servidor que permite ejecutar distintas instancias de navegador en una máquina remota, realizando varios tests en paralelo y en múltiples máquinas. Es necesario que un concentrador interactúe con las pruebas para dar acceso a las distintas instancias de navegador de forma centralizada.

Selenium fue elegido por las ventajas que ofrece a la hora de **automatizar pruebas en entornos web**, también por la **compatibilidad** con Java -eje central del proyecto-. Éste se basará en **Selenium WebDriver** y en **Selenium Client API**.

La localización de los distintos elementos web se realizará mediante la utilización del lenguaje **Xpath**, el cual se encuentra integrado en Selenium Client API.

El navegador web elegido será **Mozilla Firefox**, ya que **Selenium IDE** forma parte de sus extensiones, y será conveniente utilizar las grabaciones -que éste ofrece- en el inicio del proyecto.

3.2.2.3 MySQL

MySQL es un gestor de bases de datos relacional, multihilo y multiusuario. Provee APIs específicos que permiten acceder a dichas bases de datos. Los lenguajes de programación disponibles son: C, C++, C#, Pascal, Delphi, Eiffel, Smalltalk, **Java**, Lisp, Perl, PHP, Python, Ruby, Gambas, REALbasic, (x)Harbour (Eagle1), FreeBASIC y Tcl. También hay disponible una interfaz ODBC (MyODBC) que permite comunicarse a los lenguajes de programación -que la soportan- con MySQL.

Suele ser utilizado en aplicaciones web y en herramientas de seguimiento de errores. Tiene gran popularidad entre grandes sitios web y suele relacionarse con PHP.

Al ser un sistema de base datos relacional permite velocidad y flexibilidad, al separar los datos en distintas tablas interconectadas por relaciones definidas. Las lecturas de base de datos son muy rápidas -cuando se utiliza el motor no transaccional MyISAM- pero puede tener problemas de integridad en modificaciones de éstas en entornos muy concurridos. Suele ser importante monitorizar el rendimiento de éstas para evitar posibles problemas.

Sus principales características son:

- Amplio subconjunto del lenguaje SQL.
- Disponibilidad en distintas plataformas.
- Selección de mecanismos de almacenamiento: velocidad de operación, soporte físico, capacidad, distribución geográfica, transacciones, etc...
- Transacciones y claves foráneas.

- Seguridad.
- Replicación.
- Búsqueda e indexación de campos de texto.

La elección del gestor de base de datos MySQL, se debió a la **sencillez** de utilización, así como a su **flexibilidad** y a su **velocidad**. También por su **compatibilidad** con Java. Además, no supone ningún problema la posible falta de integridad en las modificaciones porque el sistema no será especialmente concurrente.

3.3 Implementación de la Solución Técnica

En este apartado, se describirá la implementación de cada elemento de la arquitectura: Modelo, Vista y Controlador, así como la relación entre ellos. También se añadirá una descripción con las características de la herramienta y la verificación de los requisitos del sistema.

3.3.1 Diagrama de Clases

En la siguiente figura, se muestra el diagrama de clases de la solución, en el cual puede comprobarse la **relación** existente entre los distintos módulos. Dichas clases han sido implementadas en el lenguaje de programación java.

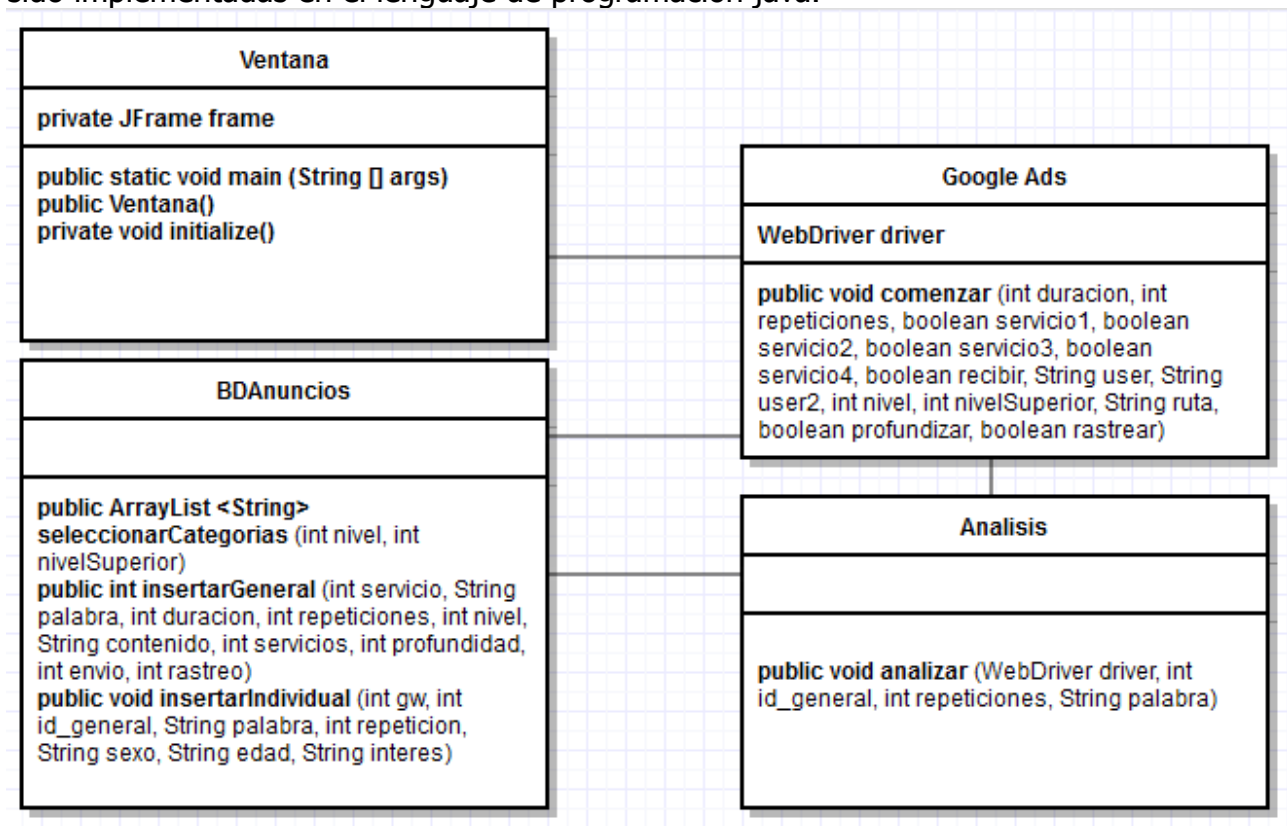


Figura 3.4: Diagrama de Clases.

3.3.2 Modelo

El modelo de la solución corresponde con la información del sistema, así como con la gestión de las peticiones de ésta. A continuación, se describe todo el diseño de la base de datos.

3.3.2.1 Diseño Base de Datos

La base de datos (**anuncios**) consta de 4 tablas, con las que se cubren todas las necesidades en cuanto a peticiones de almacenamiento y de consulta. El formato de codificación elegido es UTF-8. En las siguientes secciones, se muestra la relación entre las tablas y su descripción detallada.

Además cabe recordar que el gestor de base de datos empleado es **MySQL**, por su sencillez, velocidad, flexibilidad y concurrencia.

3.3.2.1.1 Modelo Entidad-Relación

En la siguiente figura, se muestra el modelo entidad-relación de la base de datos implementada, en el que se puede ver la **relación** existente entre las tablas:

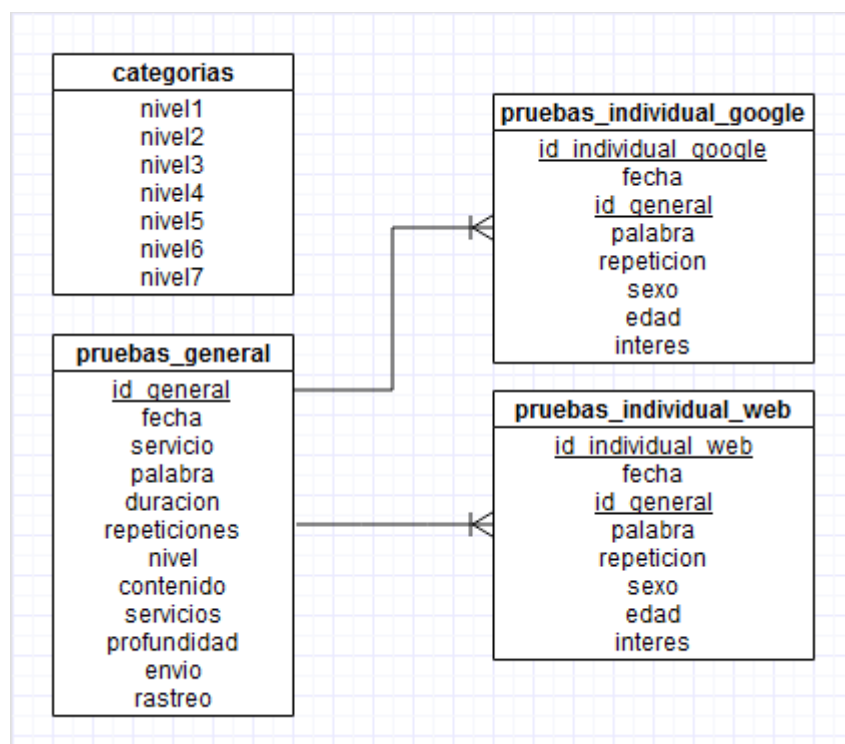


Figura 3.5: Modelo Entidad-Relación de la base de datos.

3.3.2.1.2 Descripción de Tablas

- **categorías:** Almacena todas las categorías de intereses, que Google utiliza para segmentar los anuncios basados en intereses. Pueden encontrarse en la siguiente dirección web: <https://support.google.com/ads/answer/2842480?hl=es> .

Dichas categorías de intereses tienen una distribución en forma de árbol, donde cada nivel representa las distintas subcategorías existentes. Además, previamente han sido importadas -en formato CSV- en la tabla categorías. También cabe destacar que no se van a producir modificaciones en dicha tabla.

Nombre	Tipo	Tamaño	Descripción
nivel1	varchar	29	Nivel 1 de categorías de intereses.
nivel2	varchar	61	Nivel 2 de categorías de intereses.
nivel3	varchar	64	Nivel 3 de categorías de intereses.
nivel4	varchar	64	Nivel 4 de categorías de intereses.
nivel5	varchar	57	Nivel 5 de categorías de intereses.
nivel6	varchar	52	Nivel 6 de categorías de intereses.
nivel7	varchar	26	Nivel 7 de categorías de intereses.

Tabla 3.31: Tabla categorías.

- **pruebas_general:** Almacena todos los datos generales relacionados con una prueba.

Nombre	Tipo	Tamaño	Descripción
<u>id_general</u>	int	10	Identificador general de la prueba. Clave primaria.
fecha	timestamp	6	Fecha de inicio de la prueba.
servicio	int	10	Tipo de servicio utilizado en la prueba (1: Buscador, 2: Gmail, 3: Google Maps, 4: Youtube)
palabra	varchar	200	Palabras (o asuntos) utilizados en la prueba.
duracion	int	10	Duración de cada repetición (en segundos).
repeticiones	int	10	Repeticiones de cada palabra en la prueba.
nivel	int	10	Rango de niveles -de la tabla de categorías- utilizados en la prueba. (0: no utilizado)
contenido	varchar	200	Contenido del mensaje utilizado en la prueba.
servicios	int	10	Servicios utilizados en la totalidad de la prueba.
profundidad	int	10	Profundidad de la prueba (-1: no profundizar, 0: profundizar)
envio	int	10	Envío o Recepción de mensajes de la prueba (-1: enviar, 0: enviar y recibir)
rastreo	int	10	Tipo de Rastreo de la prueba (-1: no ser rastreado, 0: ser rastreado)

Tabla 3.32: Tabla pruebas_general.

- **pruebas_individual_google:** Almacena todos los datos relevantes extraídos en una prueba para la sección de anuncios en Google.

Nombre	Tipo	Tamaño	Descripción
<u>id_individual</u> <u>_google</u>	int	10	Identificador individual de la extracción en la sección de anuncios en Google. Clave primaria.
fecha	timestamp	6	Fecha de la extracción.
<u>id_general</u>	int	10	Identificador general de la prueba. Clave foránea.
palabra	varchar	200	Palabras (o asuntos) utilizados en la prueba.
repeticion	int	10	Número de repetición de la palabra en la prueba.
sexo	varchar	200	Sexo extraído.
edad	varchar	200	Edad extraída.
interes	varchar	200	Interes extraído.

Tabla 3.33: Tabla pruebas_individual_google.

- **pruebas_individual_web**: Almacena todos los datos relevantes extraídos en una prueba para la sección de anuncios Google en la web.

Nombre	Tipo	Tamaño	Descripción
<u>id_individual</u> <u>_web</u>	int	10	Identificador individual de la extracción en la sección de anuncios Google en la web. Clave primaria.
fecha	timestamp	6	Fecha de la extracción.
<u>id_general</u>	int	10	Identificador general de la prueba. Clave foránea.
palabra	varchar	200	Palabras (o asuntos) utilizados en la prueba.
repeticion	int	10	Número de repetición de la palabra en la prueba.
sexo	varchar	200	Sexo extraído.
edad	varchar	200	Edad extraída.
interes	varchar	200	Interes extraído.

Tabla 3.34: Tabla pruebas_individual_web.

Puede observarse que las tablas pruebas_individual_google y pruebas_individual_web contienen los mismos campos. Dichas tablas podrían haberse unificado en una sola, añadiendo un campo más para identificar la sección correspondiente. Sin embargo, dicha unificación hubiera complicado en exceso el posterior análisis, con lo cual se ha decidido simplificarlo creando dos tablas diferentes.

3.3.3 Controlador

En esta sección, se explicará la implementación del controlador, el cual es el núcleo de la herramienta. Realiza la función de intermediario entre el modelo y la vista. Por lo tanto, va a intercambiar peticiones con ambos.

Dicho controlador se va a separar en tres clases: **BDAuncios**, **GoogleAds** y **Analisis**. BDAuncios será la encargada de gestionar todas las peticiones de almacenamiento y de consulta con el modelo. GoogleAds va a gestionar el desarrollo de las pruebas y la simulación del comportamiento de un usuario utilizando los servicios de Google -de manera automática-. Analisis se va a encargar de realizar la extracción inteligente de información relevante de la configuración de anuncios de Google.

Tanto GoogleAds como Analisis con la utilización de **Selenium** para automatizar la interacción con el navegador web, en este caso, **Mozilla Firefox**, y con **Xpath** para la localización de elementos HTML dinámicos.

Por lo tanto, GoogleAds va a recibir una serie de parámetros personalizados de la vista y comenzará una prueba, a su vez, iniciará la simulación del comportamiento de un usuario en un navegador web -de manera automática-. En el transcurso de prueba, realizará una serie de peticiones de selección y almacenamiento a la clase BDAuncios. Además, solicitará a la clase Analisis que haga la extracción inteligente de información relevante en los momentos adecuados, así como las peticiones de almacenamiento pertinentes a las extracciones. Finalmente, informará a la vista del estado y de las incidencias que ocurran durante la prueba. Pueden apreciarse las relaciones existentes en el diagrama de clases de la figura 3.4.

En las siguientes secciones, se describirá en profundidad la gestión de peticiones de la base de datos (clase BDAuncios), la gestión de las pruebas y la simulación del comportamiento de un usuario (clase GoogleAds) y también la extracción inteligente (clase Analisis).

3.3.3.1 Gestión Base de Datos

La gestión de las peticiones de la base de datos anuncios ha sido implementada en la clase **BDAuncios**. Ésta es la encargada de comunicarse, por un lado, con la base de datos, y por el otro, con el resto de elementos del controlador (clase GoogleAds y clase Analisis). Pueden apreciarse las relaciones existentes en el diagrama de clases de la figura 3.4.

El resto de clases del controlador hacen una serie de peticiones de consulta y de almacenamiento dirigidas a la base de datos, la clase BDAuncios recibe dichas peticiones y las envía a la base de datos -tras realizar la conexión con ella-, ésta las contesta y la clase BDAuncios envía las distintas respuestas a las peticiones solicitadas.

En la siguiente figura, se puede ver un diagrama de dicha clase:

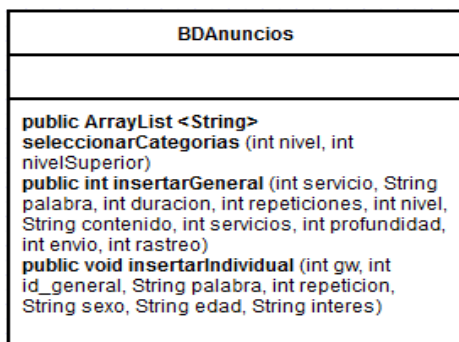


Figura 3.6: Clase BDAuncios.

A continuación, se va a explicar la funcionalidad de cada uno de los métodos de la clase:

- **Seleccionar Categorías:** El método `seleccionarCategorias` es el encargado de seleccionar un conjunto de categorías a petición de la clase `GoogleAds`.

Recibe como parámetro un nivel o un rango de niveles (nivel corresponde con el nivel seleccionado -si solo hay uno- o con el nivel inferior -en caso de ser un rango- y `nivelSuperior` corresponde con el superior de dicho rango), estos niveles corresponden con las columnas de la tabla `categorias`, la cual contiene el árbol de categorías de intereses de Google -mencionado anteriormente-.

Entonces, se hace una consulta de selección a la tabla `categorias` con dicho nivel o rango de niveles -coincidentes con las columnas de la tabla- y se devuelven las categorías seleccionadas en un `ArrayList` a la clase `GoogleAds`.

Si se produjese cualquier fallo en la lectura de la base de datos, se informaría a `GoogleAds` para que ésta lo gestionase.

- **Insertar General:** El método `insertarGeneral` es el encargado de almacenar todos los datos generales de la prueba en la tabla `pruebas_general` a petición de `GoogleAds`.

Recibe como parámetros: el servicio utilizado, la palabra utilizada (o asunto), la duración de cada repetición, las repeticiones de cada palabra, el rango de niveles seleccionado, el contenido utilizado, los servicios utilizados en la totalidad de la prueba, el tipo de profundidad, el envío o recepción de correos electrónicos y el tipo de rastreo.

Almacena en la tabla `pruebas_general` todos los parámetros recibidos, así como la fecha de almacenamiento. Finalmente, selecciona y devuelve el `id_general` del registro insertado para futuras peticiones de almacenamiento de las tablas `pruebas_individual_google` y `pruebas_individual_web` -donde el `id_general` es la clave foránea de dichas tablas-.

Si se produjese un fallo puntual en el almacenamiento, se informaría a la clase `GoogleAds` del mismo.

- **Insertar Individual:** El método `insertarIndividual` es el encargado de almacenar todos los datos relacionados con la extracción de información relevante de la configuración de anuncios de Google, en la tabla `pruebas_individual_google` -si pertenece a la sección de anuncios en Google- o en la tabla `pruebas_individual_web` -si pertenece a la sección de anuncios Google en la web-.

Recibe como parámetros: la sección en la que ha sido extraída la información (0- anuncios en Google, 1- anuncios Google en la web), el `id_general` de la prueba de la tabla

pruebas_general, la palabra utilizada, el número de repetición de la palabra, el sexo del usuario, la edad del usuario y el interés del usuario con respecto a los anuncios basados en intereses.

Se almacenan todos los parámetros recibidos y la fecha de la extracción en la tabla indicada. En este caso, la petición no devuelve nada.

Si se produjese un fallo puntual en el almacenamiento, se informaría a la clase Analisis del mismo.

3.3.3.2 Gestión Pruebas y Simulación Usuario

La clase **GoogleAds** concentra la gran parte de la lógica de negocio. En la siguiente figura, se puede ver un diagrama de dicha clase:

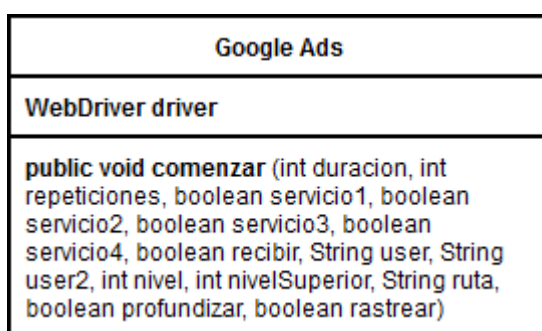


Figura 3.7: Clase GoogleAds.

Como atributo tiene un objeto de la clase **WebDriver** -facilitada por el Selenium Client API-, correspondiente al driver necesario para crear una instancia de navegador web en el que se va a realizar la prueba, y a la vez, simular el comportamiento de un usuario.

El método **comenzar** es el encargado de gestionar las pruebas y de simular el comportamiento de un usuario de manera automática. También de intercambiar peticiones, tanto con **BDAnuncios** y **Analisis** como con la vista, e informar a ésta del estado y de las incidencias ocurridas en el transcurso de la simulación.

Recibe como parámetros: la duración de cada repetición, el número de repeticiones por palabra, un booleano para el uso de cada servicio, un booleano para enviar o recibir correos, la cuenta de usuario origen y la cuenta de usuario destino, el nivel inferior y el nivel superior de la tabla de categorías, la ruta completa del fichero de entrada, un booleano para profundizar en la prueba y otro booleano para ser rastreados.

Si se produjese cualquier incidencia, la prueba finalizaría y se informaría a la vista de este hecho.

A continuación, se muestra un diagrama de flujo del método **comenzar**, para comprender mejor el funcionamiento del mismo y la gestión de las pruebas.

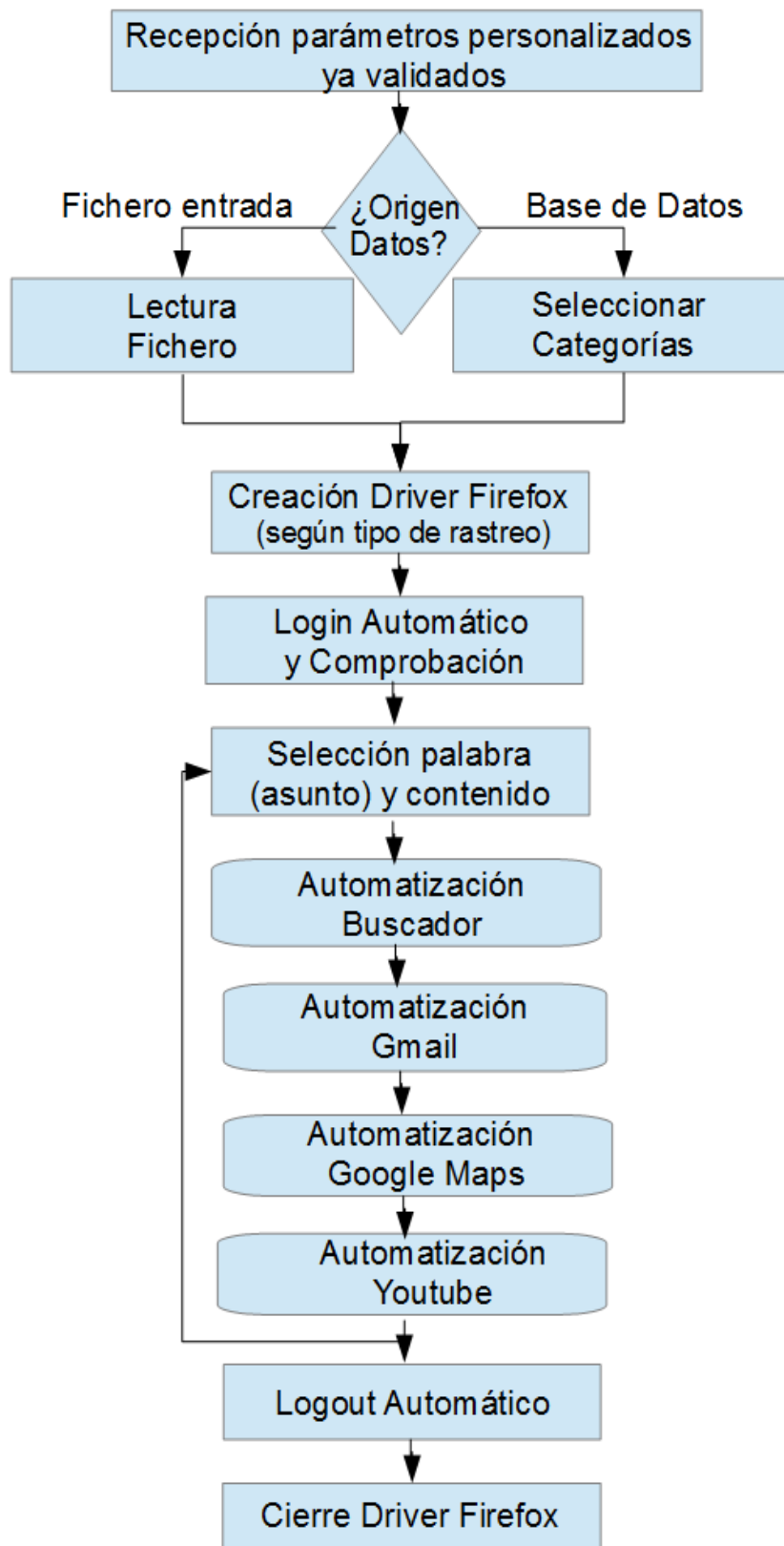


Figura 3.8: Diagrama de flujo de la gestión de pruebas. Método comenzar.

El funcionamiento es el siguiente: recepción de distintos parámetros de entrada -ya validados-, selección de palabras (asuntos) y contenidos según el origen, creación del driver, login en una cuenta, selección de palabra (asunto) y contenido, utilización de los servicios con la palabra (asunto) y el contenido elegidos -según el número de repeticiones-, y así sucesivamente, hasta finalizar el número de palabras. Finalmente, logout de la cuenta y cierre del driver.

Como se puede apreciar en el diagrama flujo anterior, hay varios bloques automatizados para simular el comportamiento de un usuario. Estos son: login, logout, buscador, gmail, google maps y youtube. Todos ellos utilizan **Selenium WebDriver** para comunicarse con el navegador web **Mozilla Firefox**, mediante los métodos que proporciona el **Selenium Client API**. Entre los métodos proporcionados, hay varios para localizar mediante **Xpath** diversos elementos HTML en páginas web.

Esta última parte ha sido una de las más difíciles de implementar, ya que no siempre las expresiones de localización eran válidas porque las páginas web -pertenecientes a Google- son bastante dinámicas y los elementos web cambian de lugar. En la mayoría de ocasiones, no tienen campos id o name fijos. Por lo que se hacía imprescindible acceder a estos elementos a través del lenguaje Xpath y con expresiones lo más robustas posibles.

También comentar que a la hora de crear el driver Firefox (FirefoxDriver) -con el que realizar la simulación-, a éste se le pasa como parámetro un perfil (FirefoxProfile) en el que podemos modificar distintos valores según nuestra conveniencia. En nuestro caso, modificar los valores correspondientes al rastreo del navegador con las opciones: `privacy.donottrackheader.enabled` y `privacy.donottrackheader.value`.

A continuación, se va a detallar un poco más en profundidad la automatización de cada uno de los servicios de Google -centro de estudio de este trabajo-:

- **Automatización Buscador (Google Maps y Youtube):** Se comienza haciendo una petición de inserción con los datos generales de la prueba en la tabla correspondiente, se realiza la búsqueda de la palabra seleccionada en el buscador (o en google maps o youtube -los tres tienen un funcionamiento similar-). Posteriormente, se interactúa con el primer resultado si se ha seleccionado profundizar en la prueba, se espera el tiempo estipulado por repetición y se hace una petición para analizar los datos de la configuración de anuncios de Google. Y así sucesivamente, hasta finalizar el número de repeticiones por palabra. En la figura 3.9, se puede ver un diagrama de flujo de estas automatizaciones.

- **Automatización Gmail:** Se comienza haciendo una petición de inserción con los datos generales de la prueba en la tabla correspondiente, se selecciona el destinatario -en función de si se ha elegido enviar o recibir-, se añade el asunto y el contenido -si se ha seleccionado profundizar en la prueba-. Entonces, se envía el mensaje, se espera el tiempo por repetición y se hace una petición para analizar los datos de la configuración de anuncios de Google. Y así sucesivamente, hasta finalizar el número de repeticiones por palabra. En la figura 3.10, se puede ver un diagrama de flujo de la automatización.

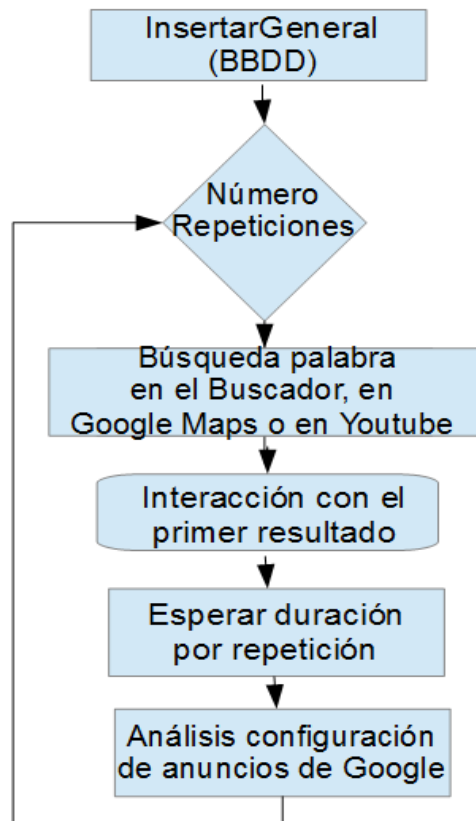


Figura 3.9: Diagrama de flujo de la automatización del Buscador, de Google Maps y de Youtube.

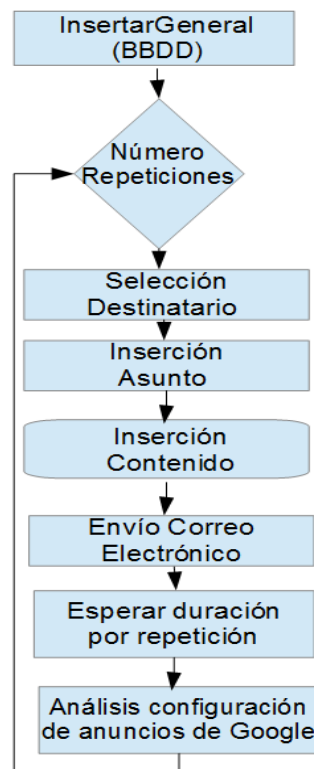


Figura 3.10: Diagrama de flujo de la automatización de Gmail.

3.3.3.3 Extracción Inteligente

La clase **Analisis** es la encargada de realizar la extracción inteligente de información relevante en la configuración de anuncios de Google. Junto con GoogleAds y BDAnuncios forman el controlador del sistema. En la siguiente figura, se puede ver un diagrama de dicha clase:

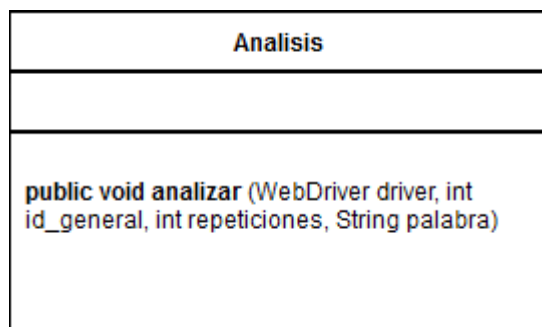


Figura 3.11: Clase Analisis.

Dicha clase recibe peticiones de análisis de la clase GoogleAds en el transcurso de las pruebas en un navegador web, y a su vez, se comunica con la clase BDAnuncios realizando peticiones de almacenamiento con la información extraída.

El método **analizar** es el encargado de extraer los valores de sexo, edad e intereses de la configuración de anuncios de Google -de manera automática- y realizar peticiones de almacenamiento correspondientes a dicha información.

Recibe como parámetros: la instancia del navegador que se estaba utilizando para la simulación, el id_general de la prueba en cuestión, el número de repetición por palabra y la palabra empleada.

La extracción está automatizada con **Selenium WebDriver** mediante el empleo de los métodos de **Selenium Client API**. Además, se utiliza **Xpath** para la localización de los distintos elementos HTML.

Esta extracción resulta bastante complicada, ya que los elementos se desplazan dentro de la propia página, en función de si hay información en la sección de anuncios Google en la web o no.

Una vez extraído cada valor de sexo, edad e interés, se realiza una petición de almacenamiento a BDAnuncios, diferenciando si proviene de la sección de anuncios en Google o de la sección de anuncios Google en la web.

Si se produjese un fallo puntual en el almacenamiento, se informaría a la clase GoogleAds del mismo.

3.3.4 Vista

En esta sección, se explicará la implementación de la vista, que consiste en una interfaz gráfica de usuario de una aplicación de escritorio -realizada con la biblioteca gráfica **Swing** de Java-.

La vista va a generar peticiones al controlador con una serie de parámetros -introducidos por el usuario- para personalizar las pruebas según sus necesidades. El controlador ejecutará dichas pruebas e informará a la vista del estado de ésta y de sus posibles incidencias, y ésta a su vez informará al usuario. Pueden apreciarse las relaciones existentes en el diagrama de clases de la figura 3.4.

La vista también guiará al usuario en la elección de las diferentes opciones con una serie de valores por defecto, así como con un breve manual de usuario sobre el uso y la configuración de la herramienta para sacar el máximo beneficio a las pruebas.

La implementación de la interfaz de usuario se realiza en la clase **Ventana**. En la siguiente figura, se puede ver un diagrama de dicha clase:

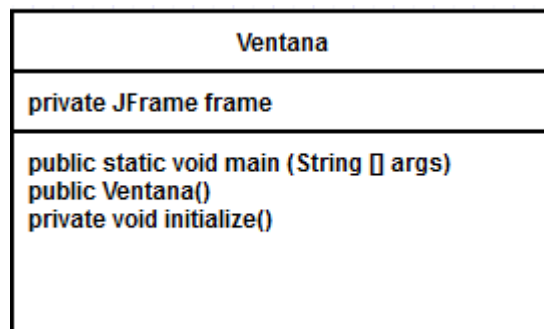


Figura 3.12: Clase Ventana.

Tiene como atributo un **JFrame** que corresponde con la instancia de la ventana principal de la aplicación.

El método **main** lanza la aplicación y crea una nueva ventana con el constructor **Ventana**, éste a su vez llama al método **initialize**.

El método **initialize** es el principal de la clase y el encargado de: inicializar la ventana con los atributos requeridos, añadir y gestionar los eventos de los distintos elementos, validar los valores introducidos por el usuario, gestionar los errores, realizar la petición de comienzo de prueba al controlador -junto con el envío de todos los parámetros seleccionados- e informar al usuario sobre el estado y las incidencias en el transcurso de ésta.

A continuación, se muestra una imagen de la interfaz gráfica de usuario implementada.

Figura 3.13: Interfaz Gráfica de Usuario.

Los parámetros a seleccionar por el usuario -en la configuración de la pruebas- son los siguientes:

- **Duración:** Duración de cada iteración en segundos. Debe ser un número entero y positivo.
- **Iteraciones:** Número de iteraciones por palabra (o palabras). Debe ser un número entero y positivo.
- **Servicios:** El usuario puede seleccionar los diferentes servicios de Google (Buscador, Gmail, Google Maps y Youtube). Debe seleccionar al menos un servicio.
- **Origen:** Cuenta origen de Google, que deseemos utilizar para realizar el estudio de los distintos servicios. Debe contener al menos un punto y una arroba. Además, por seguridad, debe ser una de las cuentas creadas específicamente para esta tarea.
- **Recibir Correos:** La opción estará habilitada si previamente hemos seleccionado el uso de Gmail. Y al desmarcarla, se habilitará el campo destino, ya que no queremos recibir correos sino mandarlos.
- **Destino:** Cuenta destino a la que queramos enviar los correos en las pruebas de Gmail. Se habilitará al desmarcar la opción recibir correos. Debe contener al menos un punto y una arroba. Además, por seguridad, debe ser una de las cuentas creadas

específicamente para esta tarea.

- **Origen de Datos:** Selecciona el origen de las palabras (asuntos) y contenidos a utilizar en la realización de la prueba. Hay dos opciones: seleccionar las palabras del árbol de categorías de intereses de Google o seleccionarlas de un fichero externo elegido por el usuario. Solo puede escogerse una de las dos opciones, si una está habilitada, la otra no lo estará, y al revés.

- **Nivel:** Nivel de la tabla de categorías de intereses de Google. Puede ser el único nivel elegido o el nivel inferior si se trata de un rango de niveles.

- **Hasta nivel:** Al marcar la opción, se habilitará la elección del nivel superior del rango.

- **Seleccionar:** Al pulsar el botón, se abre una ventana en la que se puede elegir un fichero de texto -de nuestro sistema operativo- con el que realizar la prueba. Una vez seleccionado, se mostrará el nombre del mismo al lado del botón 'Seleccionar'.

En cuanto al fichero de entrada, puede contener en cada línea una oración o dos separadas por un punto y coma. En el caso de Gmail, supondremos que la primera oración se referirá al asunto del mensaje y la segunda oración al contenido del mismo. Para el resto de servicios, la segunda oración será ignorada.

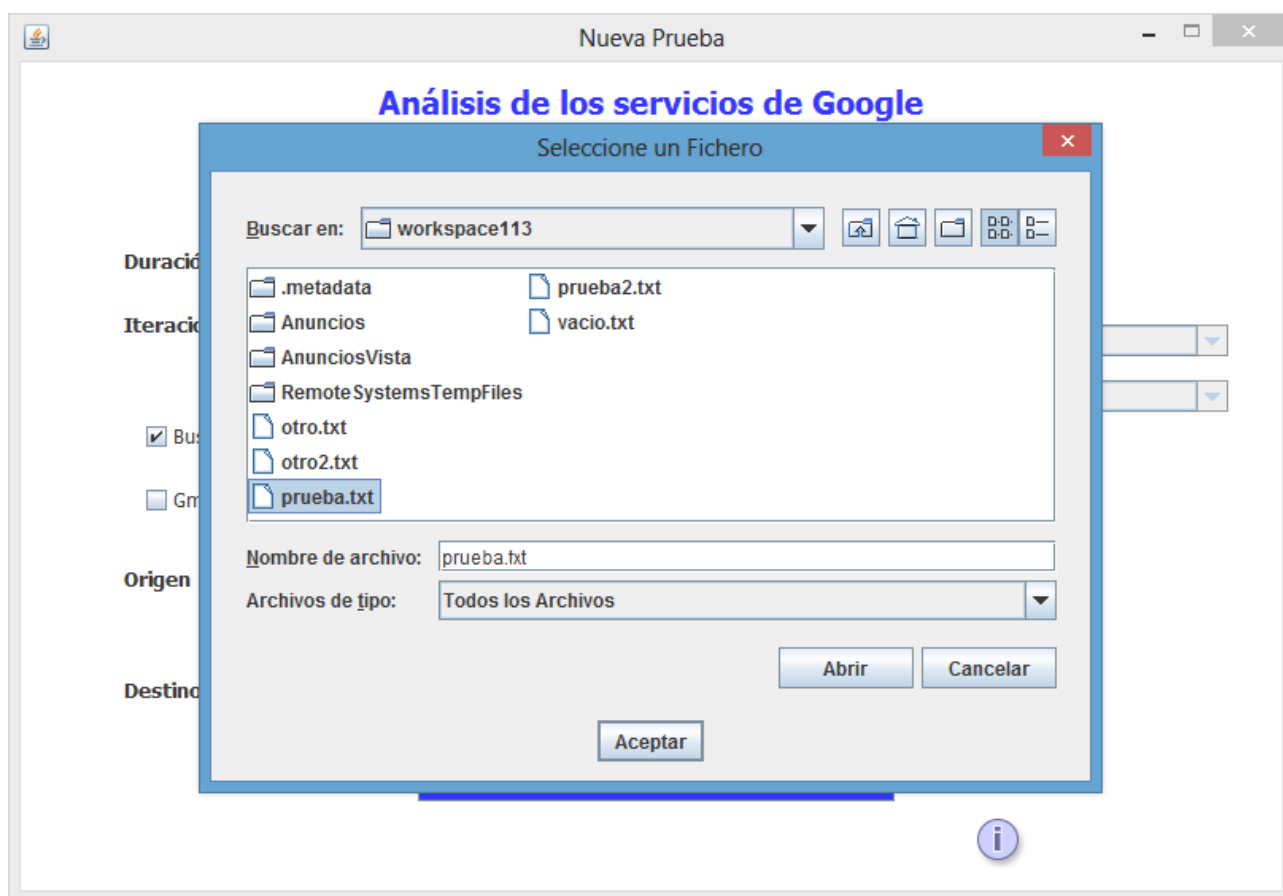


Figura 3.14: Selección Fichero de Entrada en la Interfaz Gráfica de Usuario.

- **Profundizar Prueba:** La opción se refiere a si solo queremos usar los servicios o también queremos profundizar, interactuando con los mismos: haciendo clics en los resultados de búsquedas a páginas externas, reproduciendo vídeos o rellenando el contenido de los mensajes.

- **Ser Rastreado:** Sirve para indicar a los sitios si queremos ser rastreados durante la navegación.

- **Comenzar Prueba:** Al pulsar el botón, se extraen todos los valores introducidos por el usuario y se pasan al controlador para que éste comience una nueva prueba. En el momento de pulsar el botón, se realiza la validación de todos los valores introducidos por el usuario, así como la comprobación de si se ha seleccionado un fichero.

En caso de existir cualquier incidencia en la validación o en el trascurso de la prueba, se mostrará un mensaje informativo en color rojo, informando al usuario del incidente. En caso contrario, se mostrará un mensaje informativo al usuario en color verde, informando de la finalización de la prueba.

Una vez finalizada o parada la prueba, el usuario puede volver a configurar la herramienta según sus necesidades y comenzar una nueva prueba.

Figura 3.15: Error en la validación de los datos introducidos en la Interfaz Gráfica de Usuario.

Nueva Prueba

Análisis de los servicios de Google como herramienta de marketing

Duración segundos

Iteraciones

Servicios

☒ Buscador ☐ Google Maps

☐ Gmail ☐ Youtube

Origen

☐ Recibir Correos

Destino

Origen Datos

☐ Intereses Google

Nivel

☐ Hasta nivel

☒ Fichero Externo

Seleccionar prueba2.txt

☐ Profundizar Prueba

☒ Ser Rastreado

Comenzar Prueba

Prueba no finalizada

i

Figura 3.16: Incidencia ocurrida durante la prueba.

Nueva Prueba

Análisis de los servicios de Google como herramienta de marketing

Duración segundos

Iteraciones

Servicios

☒ Buscador ☐ Google Maps

☐ Gmail ☐ Youtube

Origen

☐ Recibir Correos

Destino

Origen Datos

☐ Intereses Google

Nivel

☐ Hasta nivel

☒ Fichero Externo

Seleccionar prueba2.txt

☐ Profundizar Prueba

☒ Ser Rastreado

Comenzar Prueba

Prueba Finalizada con Éxito

i

Figura 3.17: Prueba Finalizada con Éxito.

- **Botón de información:** Al pulsarlo se abre una ventana que contiene un breve manual de usuario para orientar a éste en la configuración de la herramienta antes del comienzo de la prueba. En la figura 3.18 puede verse el contenido del mismo.

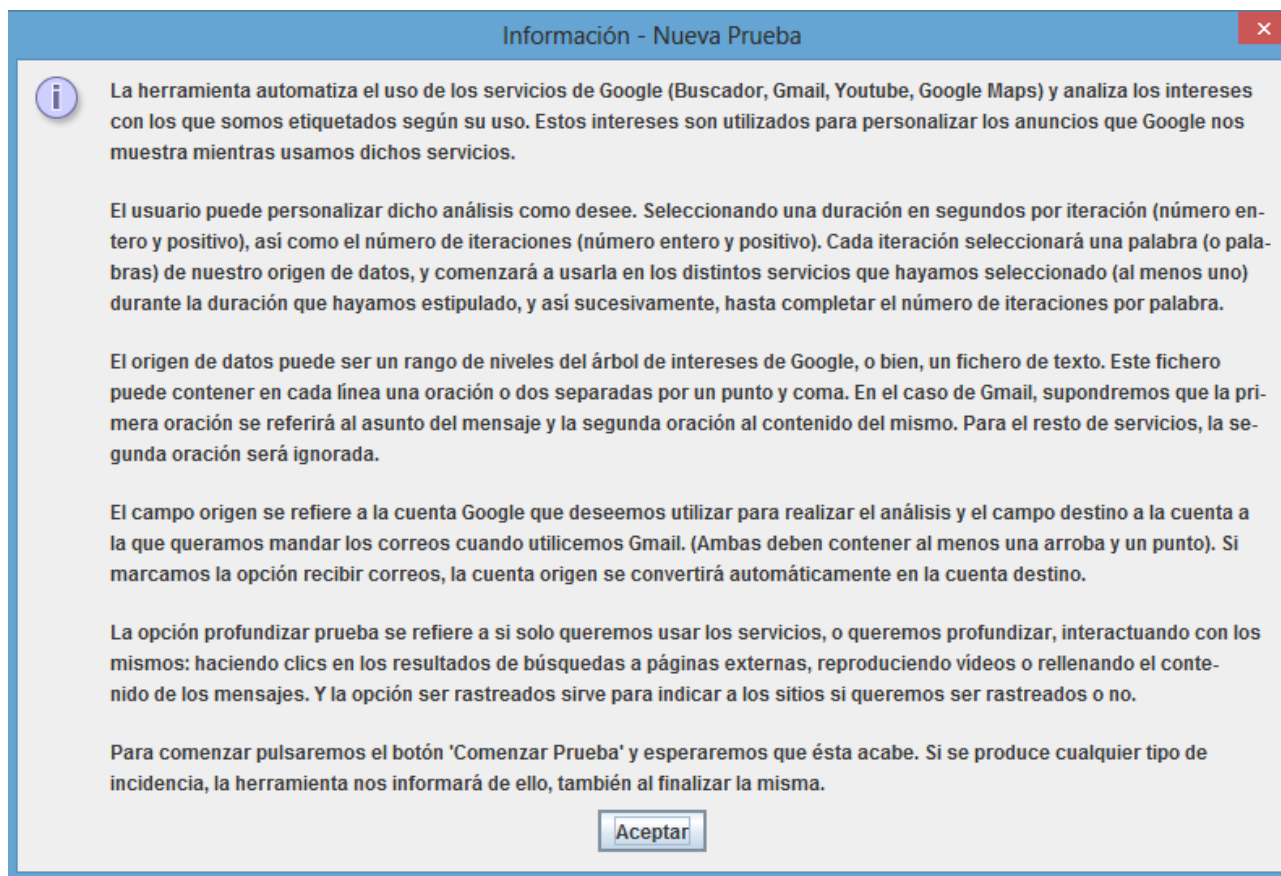


Figura 3.18: Manual de Usuario en la Interfaz Gráfica de Usuario.

Tras la descripción de la interfaz gráfica de usuario, puede apreciarse que se ha intentado realizar lo más **atractiva** posible -pese a las limitaciones que Swing ofrece-, también lo más **intuitiva** posible -para facilitar al usuario la personalización de las pruebas a realizar con los servicios de Google-. Además, se han intentado seguir -en la medida de lo posible- los **criterios** de accesibilidad, usabilidad y navegabilidad.

3.3.5 Características de la Herramienta

En este apartado, se van a describir las principales características de la herramienta implementada. Y a la vez, comprobar si se han cumplido los requisitos del sistema desarrollado.

- **Flexibilidad:** Las posibilidades que ofrece la herramienta -en cuanto a personalización de pruebas- son muy amplias. Permitiendo profundizar en muy diversos aspectos de la segmentación de intereses realizada por Google y permitiendo configurar

una infinidad de pruebas diferentes.

- **Portabilidad:** La herramienta está implementada -fundamentalmente- en lenguaje java, lo que permite que sea ejecutada en cualquier plataforma que disponga de una máquina virtual de este lenguaje.

Con la máquina virtual, una base de datos local de MySQL, conexión a Internet y un navegador Mozilla Firefox instalado, cualquiera puede utilizar la herramienta.

- **Concurrencia:** Existe la posibilidad de lanzar varias pruebas a la vez, dentro de las limitaciones de MySQL -en entornos concurrentes-. Pero hay que tener en cuenta que cuanto mayor concurrencia exista, menor será el rendimiento general del sistema.

- **Robustez:** La gestión de errores se ha llevado a cabo de manera exhaustiva, tratando de evitar cualquier posible incidente que pudiera paralizar la realización de una prueba, consiguiendo que la totalidad de la herramienta sea lo más robusta posible (incluida la localización de elementos web dinámicos).

- **Seguridad:** La aplicación es totalmente segura dentro del entorno controlado. Lamentablemente, no se puede garantizar dicha seguridad al visitar sitios web externos durante la realización de las pruebas.

- **Velocidad:** Es uno de los puntos débiles de la herramienta, que se ha intentado solventar lo mejor posible. La lentitud se produce durante la localización de los elementos web dinámicos, que en ocasiones resulta complicada. Las pruebas -cuya duración por palabra sea mayor de 20 segundos- no notarán ningún efecto.

- **Escalabilidad:** La incorporación de nuevas funcionalidades se puede hacer fácilmente, ya que la arquitectura basada en el modelo-vista-controlador permite la separación de conceptos, la reutilización de código y la escalabilidad.

En la siguiente dirección web, se puede visualizar un vídeo que contiene un **ejemplo** completo del **funcionamiento** de la herramienta desarrollada para los servicios básicos de Google, y además, se pueden verificar los requisitos del sistema:
<https://youtu.be/Gb6kjNIDatQ>

4. RESULTADOS Y EVALUACIÓN

4.1 Descripción de la Evaluación

El objetivo de la presente evaluación es analizar los distintos servicios de Google (Buscador, Gmail, Google Maps y Youtube) como herramienta de marketing. Para ello, realizaremos la **simulación** del comportamiento de un usuario -al utilizar dichos servicios- con la herramienta descrita en el capítulo anterior y analizaremos los resultados obtenidos en la configuración de anuncios de Google (Figura 2.19), en función de las opciones seleccionadas en la interfaz gráfica de usuario para llevar a cabo la prueba en cuestión.

En primer lugar, se ha seleccionado un set de 50 palabras (**set 1**) pertenecientes al nivel 2 del árbol de categorías de intereses de Google. (Para más información: <https://support.google.com/ads/answer/2842480?hl=es>). Al simular el comportamiento de un usuario -con la utilización de cada una de estas palabras-, se pretende conocer cuánto tiempo tarda Google en añadir nuevos intereses y de qué tipo son en cada uno de los servicios -partiendo de perfiles completamente vacíos, a excepción del sexo y de la edad que son obligatorios a la hora de crear la cuenta de usuario-. El análisis se va a centrar en la sección de **anuncios en Google** -pero como ya se comentó en el diseño de la solución-, ésta también almacena los datos relativos a la sección de anuncios Google en la web, lo que nos permitirá poder comparar los resultados en ambas secciones brevemente.

En segundo lugar, se ha seleccionado un set de 15 palabras (**set 2**) pertenecientes a las **categorías sensibles**, las cuales hemos obtenido tanto del árbol de categorías de intereses como de las palabras clave -que Google ofrece a los anunciantes en su servicio de marketing Google Adwords (<https://support.google.com/adwords/answer/156178?hl=es>)-. Los árboles de categorías de intereses de anunciantes y usuarios son distintos, lo que nos lleva a pensar que si Google ofrece estas categorías a sus anunciantes -para crear campañas de anuncios- puede que vulnere la privacidad de los usuarios en relación a las categorías sensibles. Al igual que en el caso anterior, se simulará el comportamiento de un usuario utilizando estas palabras de categorías sensibles y se comprobará si Google cumple con el tratamiento que debe dar a este tipo de información personal, el cual garantiza en su política de privacidad.

La configuración de la interfaz gráfica de usuario (Figura 3.13) -para realizar las pruebas comentadas- es la siguiente:

Campo	Valor
Duración	300 segundos
Iteraciones	96
Servicio	A seleccionar
Origen	Cuenta de Prueba 1

Recibir Correos	Sí
Destino	Cuenta de Prueba 2
Fichero Externo	Set 1 Set 2
Profundizar Prueba	Sí
Ser Rastreado	Sí

Tabla 4.1. Configuración de Pruebas.

Cabe destacar que habilitar la opción profundizar prueba es imprescindible para estudiar la sección de anuncios en Google, ya que es necesaria la interacción con los resultados que ofrecen los distintos servicios para que Google añada nuevos intereses en dicha sección.

La herramienta creada ofrece una infinidad de configuraciones y la seleccionada abarca solo una pequeña parte de todas las posibilidades que la herramienta ofrece. En futuras investigaciones, podrían explorarse el resto de posibilidades.

En las siguientes secciones, se muestran los resultados obtenidos con estas pruebas en los diferentes servicios, así como la comparación entre ellos.

4.2 Resultados en el Buscador

Palabra	Tiempo (minutos)	Intereses
Aviación Deportiva	115	Gimnasios
Paintball	25	Béisbol
Bebidas		
Cocina y recetas	75	Músicas del mundo
Productos y servicios para animales		
Vida salvaje	25	Nintendo
Humor	70	Veterinarios
Televisión y vídeo	180	Animales de compañía
Compra de vehículos		
Marcas de vehículos		
Cuidados faciales y corporales		
Fitness		
Servicios domésticos	165	Pop latino
Sistemas de climatización	185	Películas, Televisión y vídeo
Física	180	Películas
Matemáticas		
Grandes distribuidores y centros comerciales		
Ropa		
Agregación de feeds y marcadores sociales	75	Aplicaciones Android
Mundos virtuales		
Productos deportivos		

Resultados y estadísticas deportivos	
Educación	
Empleo	
Impresión y publicación	
Servicios para empresas	
Inversión	
Seguros	
Artículos para mujeres	125 Actividades al aire libre, Administrador de contenido, Asistencia técnica, Correo electrónico y mensajes, Internet y telecomunicaciones, Música Reggae y caribeña, Programas afiliados, Redes sociales, Servicios web, Servicios y recursos del blog, Sistemas operativos
Familia y relaciones	
Actualidad deportiva	
Cotilleos y diarios sensacionalistas	
Artículos electrónicos de consumo	
CAD y CAM	
Motores de búsqueda	
Portales web	
Juegos de ordenador y videojuegos	
Puzzles y rompecabezas	25 Perfumes y fragancias
Instituciones gubernamentales	
Jurídico	
Poesía	
Revistas	
Agencias inmobiliarias	
Casas de vacaciones	
Directorios y listados	
Referencia general	
Asia	
Europa	
Hoteles y alojamiento	
Vuelos	

Tabla 4.2: Resultados en el Buscador.

Como se puede apreciar en la tabla 4.2, solo el **24%** de las **palabras seleccionadas** (12/50) provocó que Google añadiera nuevas categorías en la sección de anuncios en Google. El tiempo medio de categorización es de **104 minutos** -en relación a los datos disponibles-. Semánticamente, el **33%** (4/12) de los **intereses añadidos** está relacionado con la palabra utilizada en la prueba -al interactuar con los resultados de las búsquedas y con la posterior visita a las distintas páginas web-.

En la sección de anuncios Google en la web, el **46%** (23/50) de las **palabras**

seleccionadas provocó que Google añadiera nuevas categorías. El tiempo medio de categorización es de **43 minutos** -en relación a los datos disponibles-. Semánticamente, el **100%** (23/23) de los **intereses añadidos** tiene relación directa con la palabra utilizada en la búsqueda.

4.3 Resultados en Google Maps

Palabra	Tiempo (minutos)	Intereses
Aviación Deportiva		
Paintball		
Bebidas		
Cocina y recetas		
Productos y servicios para animales		
Vida salvaje		
Humor		
Televisión y vídeo		
Compra de vehículos		
Marcas de vehículos		
Cuidados faciales y corporales		
Fitness		
Servicios domésticos		
Sistemas de climatización		
Física		
Matemáticas		
Grandes distribuidores y centros comerciales		
Ropa		
Agregación de feeds y marcadores sociales		
Mundos virtuales		
Productos deportivos		
Resultados y estadísticas deportivos		
Educación		
Empleo		
Impresión y publicación		
Servicios para empresas		
Inversión		
Seguros		
Artículos para mujeres		
Familia y relaciones		
Actualidad deportiva		
Cotilleos y diarios sensacionalistas		
Artículos electrónicos de consumo		
CAD y CAM		
Motores de búsqueda		
Portales web		
Juegos de ordenador y videojuegos		
Puzzles y rompecabezas		

Instituciones gubernamentales	310	Compras, Cupones y descuentos, Músicas del mundo
Jurídico		
Poesía		
Revistas		
Agencias inmobiliarias		
Casas de vacaciones	245	Música dance y electrónica, Música pop
Directorios y listados		
Referencia general	5	Bollywood y cine de Asia meridional, Reality Shows
Asia		
Europa		
Hoteles y alojamiento		
Vuelos		

Tabla 4.3: Resultados en Google Maps.

Como se puede apreciar en la tabla 4.3, solo el **6%** de las **palabras seleccionadas** (3/50) provocó que Google añadiera nuevas categorías en la sección de anuncios en Google. El tiempo medio de categorización es de **186 minutos** -en relación a los datos disponibles-. Semánticamente, el **0%** (0/3) de los **intereses añadidos** está relacionado con la palabra utilizada en la prueba -al interactuar con los resultados de las búsquedas-.

En la sección de anuncios Google en la web, **no hay datos disponibles**, al no visitar ninguna página web externa a Google.

4.4 Resultados en Youtube

Palabra	Tiempo (minutos)	Intereses
Aviación Deportiva	5	Espacio aéreo y defensa, Agencias de viajes y servicios, Proveedores de servicios, Deportes de motor
Paintball	5	Artículos de higiene, Ciclismo y accesorios, Cine de acción y aventura, Música rock, Olimpiadas, Juegos de ordenador y videojuegos, Juegos, Actualidad deportiva, Caza y tiro, Vehículos todoterreno
Bebidas	5	Compras, Carne y marisco, Empresas e industrias, Gente y Sociedad, Productos de limpieza
Cocina y recetas	5	Animales y mascotas, Casa y jardín, Empresas e industrias, Finanzas, Fitness, Gente y Sociedad, Programas de televisión, Seguros

Productos y servicios para animales	5	Perfumes y fragancias
Vida salvaje	5	Animales de compañía, Puzzles y rompecabezas
Humor	5	Programas de televisión, Gente y Sociedad
Televisión y vídeo	15	Músicas del mundo
Compra de vehículos	5	Compra de vehículos, Ley y gobierno, Publicidad y marketing
Marcas de vehículos	5	Ciclismo y accesorios , Chevrolet
Cuidados faciales y corporales	10	Deportes acuáticos
Fitness	5	Deportes de equipo, Medios y referencias para juegos, Instrumentos musicales, Referencia cinematográfica
Servicios domésticos	15	Electrodomésticos grandes, Vehículos alternativos e híbridos
Sistemas de climatización	5	Informática y electrónica, Artículos electrónicos de consumo, Finanzas, Mejora del hogar
Física	5	Tecnología espacial, Física
Matemáticas	5	Educación, Ciencias
Grandes distribuidores y centros comerciales	30	Marcas de vehículos
Ropa	5	Ropa, Juguetes, Moda y estilo
Agregación de feeds y marcadores sociales	5	Comunidades online, Operaciones empresariales
Mundos virtuales	5	Comunidades online, Cómic y animación
Productos deportivos	15	Nintendo, Cocina y recetas
Resultados y estadísticas deportivos	5	Calzado, Deportes de motor, Moda y estilo
Educación	5	Música latina, Ciencias sociales
Empleo	5	Facultades y universidades, Noticias sobre tecnología
Impresión y publicación	5	Artículos electrónicos de consumo, Camisetas, Servicios documentales y de impresión, Casa y jardín, Diseño gráfico
Servicios para empresas	5	Alimentación y bebidas, Anuncios de televisión, Hoteles y alojamiento, Seguridad informática, Smartphones
Inversión	5	Electrodomésticos grandes
Seguros	5	Compras
Artículos para mujeres	5	Deportes de equipo, Educación, Juegos de ordenador y videojuegos, Manualidades, Moda y estilo, Humor

Familia y relaciones	5	Educación infantil, primaria y secundaria, Concursos y premios
Actualidad deportiva	5	Crianza de los hijos, Instrumentos musicales
Cotilleos y diarios sensacionalistas	5	Aficiones y tiempo libre, Gente y Sociedad, Asuntos sociales y apoyo, Finanzas, Humor, Moda y estilo, Proveedores de servicios
Artículos electrónicos de consumo	5	Facultades y universidades, Instituciones gubernamentales, Cine dramático, Zapatos informales
CAD y CAM	25	Impresión y publicación
Motores de búsqueda	10	Material de oficina
Portales web	10	Música indie y alternativa, Programas de Internet
Juegos de ordenador y videojuegos	5	Informática y electrónica
Puzzles y rompecabezas	10	Juegos online
Instituciones gubernamentales	10	Instituciones gubernamentales
Jurídico	5	Facultades y universidades, Formación en informática, Diseño y desarrollo web, Ciencias sociales
Poesía	5	Poesía
Revistas	5	Compras, Electrodomésticos grandes, Grandes distribuidores y centros comerciales
Agencias inmobiliarias	10	Arte y diseño visuales, Fiestas y eventos de temporada, Préstamos
Casas de vacaciones	5	Compras
Directorios y listados	15	Agencias de viajes y servicios, Aplicaciones para móviles y add-ons, Construcción y mantenimiento, Artículos para mujeres
Referencia general	5	Formación en informática, Juegos deportivos, Biología, Deportes acuáticos, Facultades y universidades, Cómic y animación, Pequeña empresa
Asia	10	Música latina
Europa	5	Acontecimientos y listados
Hoteles y alojamiento	5	Piezas y accesorios para vehículos, Compras
Vuelos	5	Espacio aéreo y defensa, Piezas y accesorios para vehículos

Tabla 4.4: Resultados en Youtube.

Como se puede apreciar en la tabla 4.4, el **100%** de las **palabras seleccionadas** (50/50) provocó que Google añadiera nuevas categorías en la sección de anuncios en

Google. El tiempo medio de categorización es de **7.5 minutos** . Semánticamente, el **80%** (40/50) de los **intereses añadidos** está relacionado con la palabra utilizada en la prueba -al interactuar con los resultados de las búsquedas y con la posterior visualización de los diferentes vídeos-.

En la sección de anuncios Google en la web, el **98%** (49/50) de las **palabras seleccionadas** provocó que Google añadiera nuevas categorías. El tiempo medio de categorización es de **7.7 minutos** -en relación a los datos disponibles-. Semánticamente, el **96%** (47/49) de los **intereses añadidos** tiene relación directa con la palabra utilizada en la búsqueda.

4.5 Resultados en Gmail

Tras la finalización de las pruebas comentadas, **ninguna de las palabras seleccionadas** provocó que Google añadiera nuevas categorías en la sección de anuncios en Google -ni en la cuenta de origen ni en la cuenta de destino-. Pese a aparecer algunos anuncios -relacionados con el contenido de la bandeja de entrada- en la parte superior de la misma.

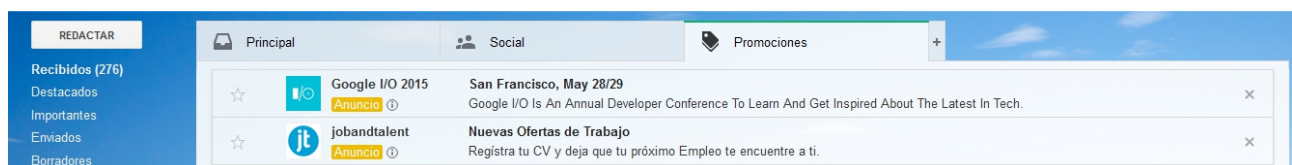


Figura 4.1: Resultados en Gmail.

En la sección de anuncios Google en la web, **no hay datos disponibles**, al no visitar ninguna página web externa a Google.

4.6 Comparativa entre Servicios

Los siguientes gráficos comparativos corresponden a los resultados -de la sección de anuncios en Google- en los diferentes servicios:

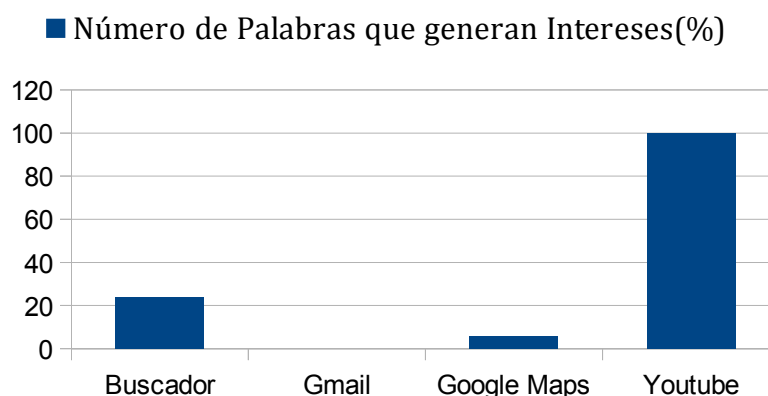


Figura 4.2. Gráfico comparativo en número de palabras que generan intereses.

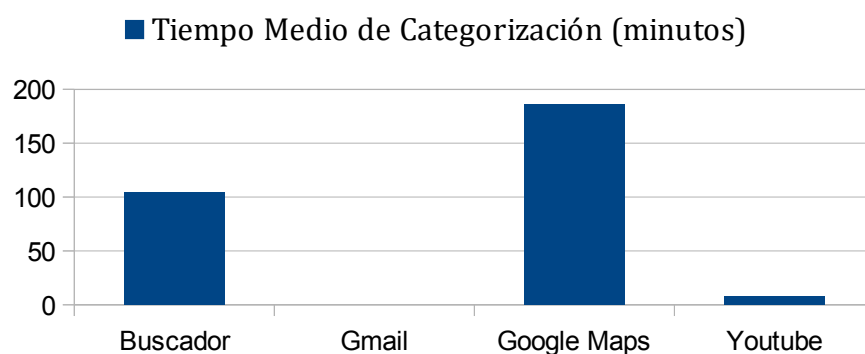


Figura 4.3. Gráfico comparativo en tiempo medio de categorización.

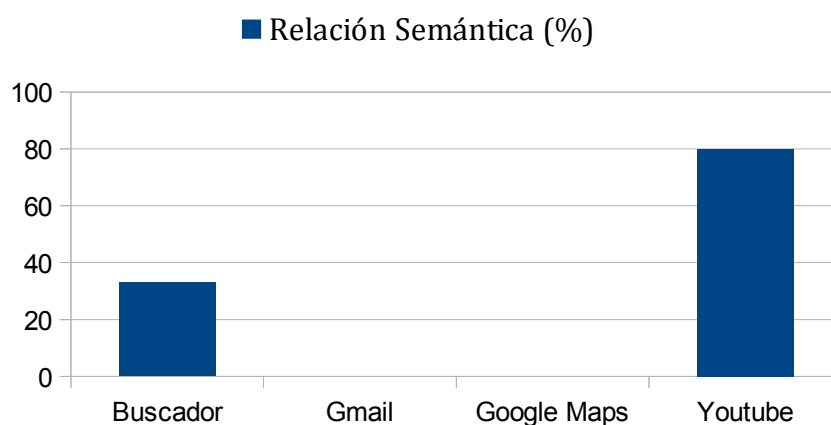


Figura 4.4. Gráfico comparativo en relación semántica.

4.7 Resultados Categorías Sensibles

Palabra	Tiempo (minutos)	Intereses
Izquierda política	5	Instituciones gubernamentales, Piezas y accesorios para vehículos, Fiestas y eventos de temporada, Préstamos
Derecha política	5	Cine clásico, Formación en informática, Urbana y Hip-Hop
Asociación sindical	5	Humor, Música latina, Programas de televisión familiares
Religión y creencias	5	Deportes de equipo, Viajes, Ciencias
Religión católica	35	Piezas y accesorios para vehículos
Religión islámica	5	Camiones y furgonetas, Humanidades, Referencia cinematográfica, Fiestas y eventos de temporada

Víctimas del 11M		
Víctimas de racismo		
Víctimas de violencia de género	45	Asuntos sociales y apoyo
Enfermedad de alzheimer	10	Referencia cinematográfica, Referencia musical
Cancer	10	Mejora del hogar, Seguridad pública, Biología
Enfermedades de transmisión sexual		
Contactos gay		
Mujeres swingers		
Sex shop		

Tabla 4.5: Resultados Categorías Sensibles.

Las pruebas de las categorías sensibles han sido realizadas en **Youtube**, ya que es el servicio que ofrece mayores prestaciones tanto en tiempo de categorización, número de palabras que provocan intereses añadidos, así como en relación semántica.

Como se puede apreciar en la tabla 4.5, solo el **60%** de las **palabras seleccionadas** (9/15) provocó que Google añadiera nuevas categorías en la sección de anuncios en Google. El tiempo medio de categorización es de **14 minutos** -en relación a los datos disponibles-. Semánticamente, el **33%** (3/9) de los **intereses añadidos** está relacionado ligeramente con la palabra utilizada en la prueba. Sin embargo, las categorías de intereses añadidas son demasiado **genéricas** como para afirmar que infringen la privacidad de los usuarios.

En la sección de anuncios Google en la web, el **20%** (3/15) de las **palabras seleccionadas** provocó que Google añadiera nuevas categorías de intereses. El tiempo medio de categorización es de **25 minutos** -en relación a los datos disponibles-. Semánticamente, el **0%** (0/3) de los **intereses añadidos** tiene relación con la palabra utilizada en la búsqueda.

Comparación Set 1 y Set 2

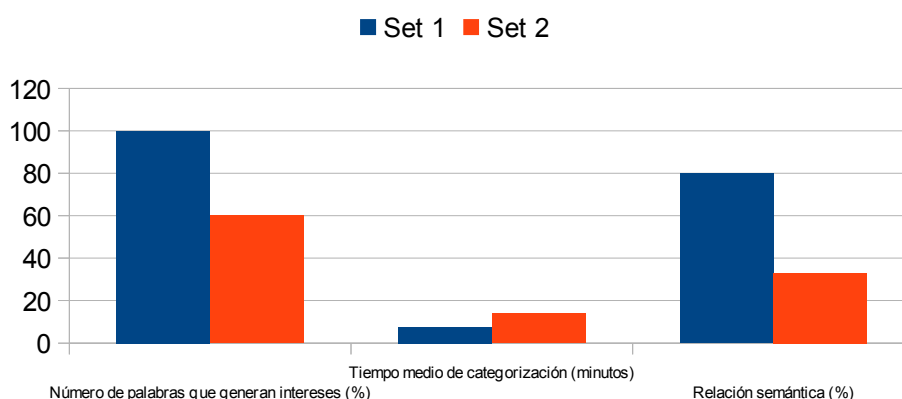


Figura 4.5. Gráfico comparativo entre set 1 y set 2.

5. PRESUPUESTO Y PLANIFICACIÓN DEL TRABAJO

5.1 Planificación del Trabajo

En este apartado, se detallarán las tareas realizadas en las distintas fases del ciclo de vida del proyecto. Además, se mostrará un diagrama de Gantt con la planificación temporal seguida.

5.1.1 Fases del Ciclo de Vida

El modelo de ciclo de vida utilizado ha sido **cascada realimentado**, elegido por su **eficacia** y **simplicidad**. Además, al contar con cierta realimentación permite la detección y corrección de errores. En la figura 5.1, se pueden apreciar las fases más habituales de un proyecto de software con este modelo de ciclo de vida.

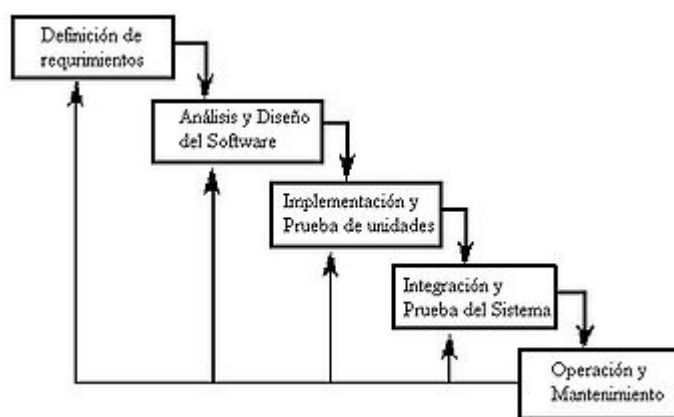


Figura 5.1: Modelo Cascada Realimentado. Fuente: es.wikipedia.org

El ciclo de vida del proyecto se ha separado en las siguientes fases:

- **Inicio:** Fase en la que se realizan una serie de estudios previos, cuyo fin es determinar los objetivos y el alcance del proyecto.
- **Análisis:** Fase en la que se realizan distintos análisis del problema, de su ámbito y de las tecnologías disponibles, cuyo fin es realizar un análisis de requisitos del problema.
- **Diseño:** Fase en la que se realiza el diseño de la arquitectura, del entorno tecnológico y del plan de pruebas.
- **Implementación:** Fase en la que se lleva a cabo el desarrollo de los distintos módulos de la arquitectura.

- **Integración y pruebas:** Fase en la que se integran los distintos módulos y se lleva a cabo el plan de pruebas. También la evaluación de los servicios de Google.

- **Cierre:** Fase en la que se extraen las conclusiones de la evaluación y se documenta todo el trabajo realizado, dando por finalizado el proyecto.

A continuación, se detallan cada una de las tareas realizadas en cada fase:

	Nombre	Duración	Inicio	Terminado	Predecesores	Nombres del Recurso
1	Inicio	15 days	1/10/14 8...	15/10/14 ...		
2	Estudio Inicial: Necesidades y Deficiencias	3 days	1/10/14 8:00	3/10/14 17...		Jefe de Proyecto
3	Análisis cobertura Necesidades y mitigación Deficiencias	4 days	4/10/14 8:00	7/10/14 17...	2	Jefe de Proyecto
4	Estudio vías alternativas	2,5 days	8/10/14 8:00	10/10/14 1...	3	Jefe de Proyecto
5	Estimación recursos, costes y plazos	2,5 days	10/10/14 1...	12/10/14 1...	4	Jefe de Proyecto
6	Determinación objetivos y alcance	3 days	13/10/14 8...	15/10/14 1...	5	Jefe de Proyecto
7	Objetivos y alcance definidos	0 days	15/10/14 1...	15/10/14 1...	6	Jefe de Proyecto
8	Análisis	37,5 days	16/10/14 ...	22/11/14 ...	7	
9	Análisis Estudio del Arte	8 days	16/10/14 8...	23/10/14 1...		Analista
10	Análisis Marco Regulatorio	3,5 days	24/10/14 8...	27/10/14 1...	9	Analista
11	Análisis Google como herramienta de marketing	7 days	27/10/14 1...	3/11/14 12...	10	Analista
12	Análisis Requisitos	7 days	3/11/14 13...	10/11/14 1...	11	Analista
13	Análisis Tecnologías	6 days	10/11/14 1...	16/11/14 1...	12	Analista
14	Planificación detallada y Presupuesto	5,5 days	16/11/14 1...	21/11/14 1...	13	Analista
15	Análisis Finalizado	0,5 days	22/11/14 8...	22/11/14 1...	14	Jefe de Proyecto
16	Diseño	37,5 days	22/11/14 ...	29/12/14 ...	15	
17	Diseño Arquitectura	11 days	22/11/14 1...	3/12/14 12...		Analista
18	Diseño Entorno Tecnológico	11 days	3/12/14 13...	14/12/14 1...	17	Analista
19	Diseño Plan de Pruebas	9 days	14/12/14 1...	23/12/14 1...	18	Analista
20	Evaluación Requisitos	3 days	23/12/14 1...	26/12/14 1...	19	Analista
21	Evaluación Diseño	3 days	26/12/14 1...	29/12/14 1...	20	Analista
22	Diseño aprobado	0,5 days	29/12/14 1...	29/12/14 1...	21	Jefe de Proyecto
23	Implementación	75 days	30/12/14 ...	14/03/15 ...	22	
24	Formación	7 days	30/12/14 8...	5/01/15 17...		Programador
25	Automatización Servicios	25 days	6/01/15 8:00	30/01/15 1...	24	Programador
26	Extracción Inteligente	15 days	31/01/15 8...	14/02/15 1...	25	Programador
27	Diseño Base de Datos	5 days	15/02/15 8...	19/02/15 1...	26	Programador
28	Gestión Base de Datos	5 days	20/02/15 8...	24/02/15 1...	27	Programador
29	Diseño Interfaz Gráfica de Usuario	5 days	25/02/15 8...	1/03/15 17...	28	Programador
30	Desarrollo Interfaz Gráfica de Usuario	5 days	2/03/15 8:00	6/03/15 17...	29	Programador
31	Integración Inicial	7,5 days	7/03/15 8:00	14/03/15 1...	30	Programador
32	Implementación finalizada	0,5 days	14/03/15 1...	14/03/15 1...	31	Analista; Jefe de Proyecto
33	Integración y Pruebas	37,5 days	15/03/15 ...	21/04/15 ...	32	
34	Validación Requisitos	3 days	15/03/15 8...	17/03/15 1...		Analista
35	Determinación de correcciones	3 days	18/03/15 8...	20/03/15 1...	34	Analista
36	Integración Finalizada	0,5 days	21/03/15 8...	21/03/15 1...	35	Jefe de Proyecto
37	Ejecución Evaluación	25 days	21/03/15 1...	15/04/15 1...	36	Programador
38	Evaluación Plan de Pruebas	5,5 days	15/04/15 1...	20/04/15 1...	37	Analista
39	Evaluación Finalizada	0,5 days	21/04/15 8...	21/04/15 1...	38	Jefe de Proyecto
40	Cierre	37,5 days	21/04/15 ...	28/05/15 ...	39	
41	Documentación	37 days	21/04/15 1...	28/05/15 1...		Programador
42	Supervisión Final	0,5 days	28/05/15 1...	28/05/15 1...	41	Analista; Jefe de Proyecto

Figura 5.2: Tareas Realizadas.

5.1.2 Diagrama de Gantt

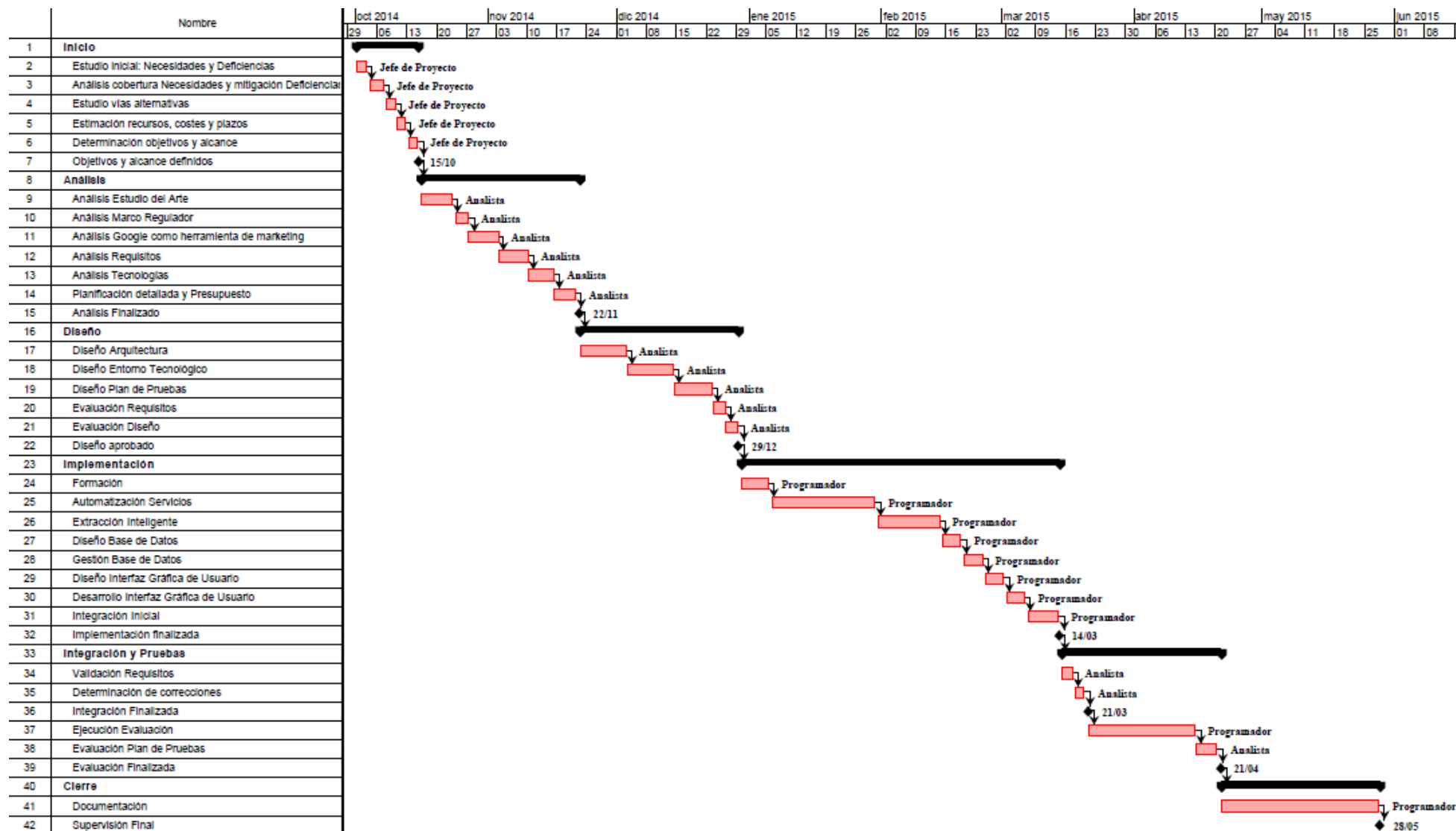


Figura 5.3: Diagrama de Gantt.

En el diagrama de Gantt, se pueden apreciar las tareas realizadas, la planificación temporal de las mismas, así como los recursos asignados.

El proyecto ha tenido una duración de 8 meses (del 1 de Octubre de 2014 al 28 de Mayo de 2015) con una media de 2 horas dedicadas por día, haciendo un total de **480 horas**.

Además, ha sido necesario realizar tres tipos de roles: **Jefe de Proyecto, Analista y Programador**.

5.2 Presupuesto del Trabajo

En este apartado, se detallará el presupuesto empleado en el desarrollo del proyecto, tanto en material como en personal.

En cuanto al presupuesto de personal, el número de horas realizadas corresponde con el número de horas de la planificación -comentada en el apartado anterior-.

Concepto	Horas	€/hora	Total
Jefe de Proyecto	36	40	1.440,00 €
Analista	173	35	6.055,00 €
Programador	271	30	8.130,00 €
Subtotal			15.625,00 €

Tabla 5.1: Subtotal Personal.

En cuanto al presupuesto de material, se ha intentado reducir al máximo los costes, utilizando la mayoría de programas de distribuciones de software libre.

Concepto	Coste	Unidades	Total
Ordenador Personal	800,00 €	1	800,00 €
Licencia Windows 8	0,00 €	1	0,00 €
Servidor Externo	1.000,00 €	1	1.000,00 €
Licencia Ubuntu	0,00 €	1	0,00 €
Licencia Máquina Virtual de Java	0,00 €	2	0,00 €
Licencia Eclipse	0,00 €	1	0,00 €
Licencia Firefox	0,00 €	2	0,00 €
Licencia Selenium IDE	0,00 €	1	0,00 €
Licencia Xampp (MySQL)	0,00 €	2	0,00 €
Licencia OpenOffice	0,00 €	1	0,00 €
Licencia ProjectLibre	0,00 €	1	0,00 €
Subtotal			1.800,00 €

Tabla 5.2: Subtotal Material.

Finalmente, el presupuesto total del proyecto es el siguiente:

Subtotal Personal	15.625,00 €
Subtotal Material	1.800,00 €
TOTAL	17.425,00 €

Tabla 5.3: Presupuesto Total.

6. CONCLUSIONS

Delving into the advertising network of Google has been an amazing but arduous task. The algorithms, used by the company to customize interest-based ads, are extremely **complex** and the number of involved factors is too big. So, this task sometimes seemed to lead us increasingly far from understanding the use of the Google services as a marketing tool.

Youtube is the service that provides better time performance, more number of words generating new interests and also better semantic relationship. Moreover, it reaches good results only in 5 minutes. Followed by the **search engine** that provides discrete results but, due to its relevance, it's suitable to consider it. Finally, there is no enough information about the use of **Google Maps** and **Gmail** as a marketing tool.

With regards to the sensitive categories, it's curious that Google offers keywords and ad interest categories -to its advertisers- that contains some sensitive categories. However, after testing the processing, it hasn't detected that the company doesn't **satisfy** the terms of its privacy policy or the applicable laws on data protection.

On the other hand, the developed tool includes **all** available **options** for the creation of interest-based profiles. Unfortunately, it has only been possible to explore a small part of all possibilities that the tool offers. In future researches, it could continue with the exploration of another offered options by the tool, or to even study another Google services. The tool could be also adapted to another web browsers, the GUI could be improved using JavaFX, or the system could be adapted as a web application, even as a mobile application using Selendroid.

Moreover, the online marketing industry is booming. It expects to continue growing along with the **Big Data** to create more personal profiles and less based on demographics. So it's necessary that the authorities speed up the approval of adapted legislation to the new needs. The use of Big Data as a marketing tool is another interesting line of research.

Finally, it can check that all objectives, that were defined at the beginning of the work, have been achieved. Personally, I'm very satisfied, I think it has been a great experience to me.

Ad astra per aspera. Seneca.

6. CONCLUSIONES

Adentrarse en las profundidades de la red de marketing de Google ha sido una tarea fascinante, y a la vez, ardua. Los algoritmos utilizados por la compañía -para la segmentación de anuncios basados en intereses- son extremadamente **complejos** y el número de factores involucrados es demasiado grande. Por lo que, en ocasiones, esa tarea parecía llevarnos cada vez más lejos de comprender el funcionamiento de los distintos servicios como herramienta de marketing.

Youtube es el servicio que presenta mejores prestaciones -tanto en número de palabras que generan nuevos intereses, en tiempo de categorización, así como en relación semántica-, alcanzando resultados bastante aceptables con solo 5 minutos de utilización. A continuación, el **buscador** presenta unos resultados discretos. Sin embargo, debido a la importancia del mismo, es conveniente tenerlo presente. Finalmente, no tenemos información suficiente sobre el funcionamiento de **Google Maps** y de **Gmail** como herramientas de marketing.

En cuanto a las categorías sensibles, es curioso que Google ofrezca palabras clave y un árbol de categorías de intereses -a sus anunciantes- que contengan categorías sensibles. Sin embargo, a priori con las pruebas realizadas, **no** hemos detectado que la compañía **incumpla** su política de privacidad ni las leyes vigentes de protección de datos.

Si bien es cierto, la solución técnica desarrollada incluye **todas** las **posibilidades** que influyen en la creación de perfiles basados en intereses. Lamentablemente, solo se ha podido explorar una mínima parte de todas las posibilidades que la herramienta ofrece. En futuras investigaciones, podría continuarse con la exploración de las distintas opciones ofrecidas o estudiar otros servicios ofrecidos por Google. Además, la herramienta podría adaptarse a otros navegadores web, la interfaz gráfica de usuario podría mejorarse con la tecnología JavaFX, incluso adaptarse como aplicación web o como aplicación móvil con la tecnología Selendroid.

Por otro lado, el sector del marketing online se encuentra en pleno auge. Se espera que éste siga creciendo junto con el **Big Data** para basarse -cada vez más- en perfiles más personales y menos basados en datos demográficos. Lo que hace imprescindible que las autoridades competentes agilicen la aprobación de una legislación adaptada a las nuevas necesidades. La utilización del Big Data como herramienta de marketing es otra línea de investigación a tener en cuenta.

Finalmente, se puede comprobar que todos los objetivos planteados al principio del trabajo se han superado con creces. Personalmente, estoy muy satisfecha, creo que ha sido una experiencia que me ha enriquecido enormemente.

Ad astra per aspera. Séneca.

GLOSARIO DE TÉRMINOS

- **Amazon:** Compañía de comercio electrónico y servicios de cloud computing.
- **Big Data:** Sistemas informáticos basados en la acumulación a gran escala de datos y de los procedimientos usados para identificar patrones recurrentes dentro de esos datos.
- **bing:** Buscador web de Microsoft.
- **Blogger:** Servicio que permite crear y publicar un blog en línea. Propiedad de Google.
- **Dropbox:** Servicio de almacenamiento de archivos multiplataforma en la nube.
- **eBay:** Sitio destinado a la subasta de productos a través de Internet.
- **Facebook:** Red Social.
- **Gmail:** Servicio de correo electrónico. Propiedad de Google.
- **Google +:** Red Social. Propiedad de Google.
- **Google Maps:** Servicio de Mapas en la red. Propiedad de Google.
- **Java:** Lenguaje de programación de propósito general, concurrente y orientado a objetos.
- **LinkedIn:** Sitio web orientado a negocios.
- **Microblog:** Servicio que permite enviar y publicar mensajes breves.
- **Mozilla Firefox:** Navegador web libre y de código abierto.
- **MySQL:** Sistema de gestión de base de datos relacional, multihilo y multiusuario.
- **Outlook:** Cliente de correo electrónico y noticias de Microsoft.
- **Paypal:** Compañía de comercio electrónico que permite pagar en sitios web.
- **Plugin:** Extensión.
- **Podcast:** Archivo de audio gratuito, que se puede descargar y reproducir.
- **Selenium:** Entorno de pruebas de software para aplicaciones basadas en la web.
- **Skype:** Servicio que permite comunicaciones de texto, voz y vídeo sobre Internet.
- **Spotify:** Aplicación empleada para la reproducción de música vía streaming.

- **Swing**: Librería gráfica de Java.
- **Tweet**: Mensaje en Twitter.
- **Twitter**: Servicio de Microblogging.
- **Whatsapp**: Aplicación de mensajería instantánea para teléfonos inteligentes.
- **Widget**: Pequeña aplicación.
- **Wiki**: Sitio web colaborativo.
- **Wordpress**: Sistema de gestión de contenidos.
- **Xpath**: Lenguaje que permite crear expresiones que recorren y procesan un documento XML.
- **Yahoo Mail**: Correo electrónico de Yahoo.
- **Youtube**: Sitio web donde los usuarios pueden subir y compartir vídeos. Propiedad de Google.

GLOSARIO DE ACRÓNIMOS

- **API**: Application Programming Interface
- **CSV**: Comma-Separated Values
- **GUI**: Graphical User Interface
- **HTML**: HyperText Markup Language
- **IDE**: Integrated Development Environment
- **JVM**: Java Virtual Machine
- **LOPD**: Ley Orgánica 15/1999 de 13 de diciembre de Protección de Datos de Carácter Personal
- **MOOC**: Massive Open Online Course
- **MVC**: Modelo-Vista-Controlador
- **PPC**: Pay Per Click
- **PPF**: Puntos Publicitarios Fraudulentos
- **SEM**: Search Engine Marketing
- **SEO**: Search Engine Optimization
- **SMS**: Short Message Service
- **SQL**: Structured Query Language
- **URL**: Uniform Resource Locator

REFERENCIAS

- [1] <http://es.wikipedia.org>
- [2] <http://en.wikipedia.org>
- [3] <http://www.rae.es>
- [4] <https://support.google.com/>
- [5] <https://www.facebook.com>
- [6] <https://twitter.com/>
- [7] <https://www.linkedin.com>
- [8] <http://docs.seleniumhq.org>
- [9] <http://www.w3schools.com>
- [10] <http://selendroid.io>
- [11] <https://www.java.com>
- [12] <https://eclipse.org/>
- [13] <https://www.mysql.com/>
- [14] <https://www.apachefriends.org>
- [15] <https://www.mozilla.org>
- [16] <http://www.a pep.es/>
- [17] <https://www.agpd.es>
- [18] <http://www.privacidadlogica.es/2015/03/12/ultimas-modificaciones-del-reglamento-europeo-de-proteccion-de-datos>
- [19] <http://www.elmundo.es/elmundo/2010/05/15/navegante/1273920877.html>
- [20] <http://www.notariallopi.es/blog/i/145/73/primera-sentencia-en-espana-contragoogle-por-proteccion-de-datos-sensibles>

APPENDIX A - EXTENDED SUMMARY

Problem Statement

Online marketing arose from the need to adapt traditional marketing to the new market needs -where millions of people from around the world daily use the network-. The Internet is completely different from previous media, it's interactive and it allows to get feedback from users. In this manner, marketers are able to customize online marketing strategies to potential customers.

In this line, the latest trends in advertising focus on the audience **segmentation according to the user profiles** on the network, and so, the ad campaigns can be shown only to those customers who can be really interested in them.

These profiles are based on web history, search history, amount of time on each page, the visited links, among others. Another factors are: geography, demographics or contextual content of the site.

The main objective is still to promote and to sell products and services. However, the strategies have completely changed. The new tasks can be: content management, public relations, online reputation management, sales and customer service.

The most popular online marketing strategies are: web design, **advertisements on websites** (banners or another formats), search engine optimization (**SEO**), search engine marketing (**SEM**), e-mail marketing, **social media marketing** and affiliate networks.

Also mobile marketing, viral marketing, blog marketing, microblogging marketing, podcasting, video marketing, widgets or wikimarketing. In addition to performance marketing -including cost per click, cost per impression, cost per lead or cost per sale- and website audit -to measure reputation, popularity and influence of the websites-.

Web navigation isn't an anonymous activity -as anyone could think-. Both Internet providers and website administrators can collect a large amount of personal information during navigation (for example, using cookies).

So the rise of the Internet has led the implantation of some international **laws on data protection**, that try to protect the **users' privacy** as much as possible, especially the data that are related to the **sensitive categories** -those that need special attention for being linked to the personal privacy, fundamental rights and civil liberties-.

The proposal for **the General Data Protection Regulation** (GDPR) was approved by the Commission, the Parliament and the Council of the European Union in early 2012. These reforms will replace the current legislation in Spain, in terms of data protection: 15/1999, 13th December, Law on protection of personal data (LOPD).

The objective of this proposal is to revise the current laws on data protection because of the fast technological development and the significant increase in data exchange. Also to unify criteria in the EU member countries. It expects that the changes become effective in **2016**.

Google is the largest collector of personal data around the world and this work is going to focus on studying and analyzing its services as a marketing tool: how our personal data are collected -using some basic services (**Search engine, Gmail, Google Maps and Youtube**)- to place interest-based ads inside its advertising network and how the sensitive categories are processed.

Moreover, Google is one of the leading companies in terms of online marketing, due to its extensive worldwide coverage (90-95%), placing Google in a privileged position compared to its competitors.

Firstly, the top positions of the results of its **search engine** are the most desired place by anyone who wants to make website search engine optimized (**SEO**).

Google is also a strategic place for those who want to pay for placing their ad campaigns (**SEM**) in one of the offered places for the company: at its services or its extensive display network. Among its products are **Google Adwords** and **Google Adsense** to create ad campaigns for a target audience.

These campaigns are also shown on social networks like **Youtube** or **Google+**, so Google also covers the area of **social media marketing**.

Most of offered services by Google are free. However, it's not completely true, the users must "pay" a price by using these services, and that price is a part of their **privacy** on the network, that Google uses to customize profitable ad campaigns.

Its **privacy policy** clearly specifies that all personal data -used in their services- are property of Google, and therefore, source of its advertising network. And the sale of the privacy is totally accepted by the users when they use those services.

The first objective of this work is the creation of a tool to **simulate** the **user behavior** and so to automatically use the Google services to know why, when and how the users are inserted into a particular market segment, according to their actions in the different services. The second objective is to check that Google really satisfies its privacy policy and the current laws on data protection for the **sensitive categories**.

All ad interest categories -used by Google to create user profiles- are in the following link: <https://support.google.com/ads/answer/2842480?hl=en> . And the **settings for Google Ads** are at the figure below:

Settings for Google Ads

Ads enable free web services and content. These settings control the types of ads you see.





	Ads on Google	Google Ads Across the Web [?]
	<div> Search</div> <div> Gmail</div> <div> YouTube</div>	<div></div> <div>Google Ads Across the Web</div>
Gender	Unknown Visit your Google Profile	Male Edit <small>Based on the websites you've visited</small>
Age	Unknown Visit your Google Profile	25-34 Edit <small>Based on the websites you've visited</small>
Interests	Unknown Edit <small>From your previous searches</small>	Autos & Vehicles, and 7 more. Edit <small>Based on the websites you've visited</small>

Figure A.1: Settings for Google Ads.

Technical Design Solution

The software architectural pattern -for the solution- is the **Model-View-Controller** (MVC). This kind of architecture separates the data (model), the business logic (controller) and the graphical user interface (view) in different modules, facilitating the separation of concepts, the code reuse and the scalability of the system for future improvements and extensions. So the modules of the chosen solution are the followings:

- **View:** It's the data and the business logic in a suitable format so that the user can interact to the system. In our case, it will be a graphical user interface (GUI) in a desktop application.

The user will interact to the system -customizing the test cases according to the needs-. And the system will interact to the user -reporting on the status of the test cases or any happened incident-. Moreover, some defaults options will guide to the user -during the customization- and also a section of information about the operation of the tool.

Java will be the main technology in the graphical user interface, specifically, the **Swing** library. The use of java will provide portability to the system and the tool could be used on any platform.

Figure A.2: Graphical User Interface.

- **Controller:** It receives requests from the view for customizing the test cases and it makes requests to the model for managing data. It can also send information from the model to the view. It's the intermediary between the view and the model.

In our case, it will be responsible for simulating the user behavior in the different Google services. Its main functions are: automatically management of the requested test cases by the user in the view, intelligent system for data extraction in settings for Google ads, storage requests to the model -general and extracted data from the test cases-, also requests for information to the model -about ad interest categories provided by Google-. It will also report on the status of the test cases, or any happened incident, to the view.

Java will be the main technology in the controller. **Selenium WebDriver** will be used for automating the test cases and for the relevant data extraction. And **Xpath** will be used to locate dynamic HTML elements in the different websites -property of Google-. Both supported by Java. Finally, **Mozilla Firefox** will be the chosen web browser for the simulation of the user behavior.

SYSTEM ARCHITECTURE

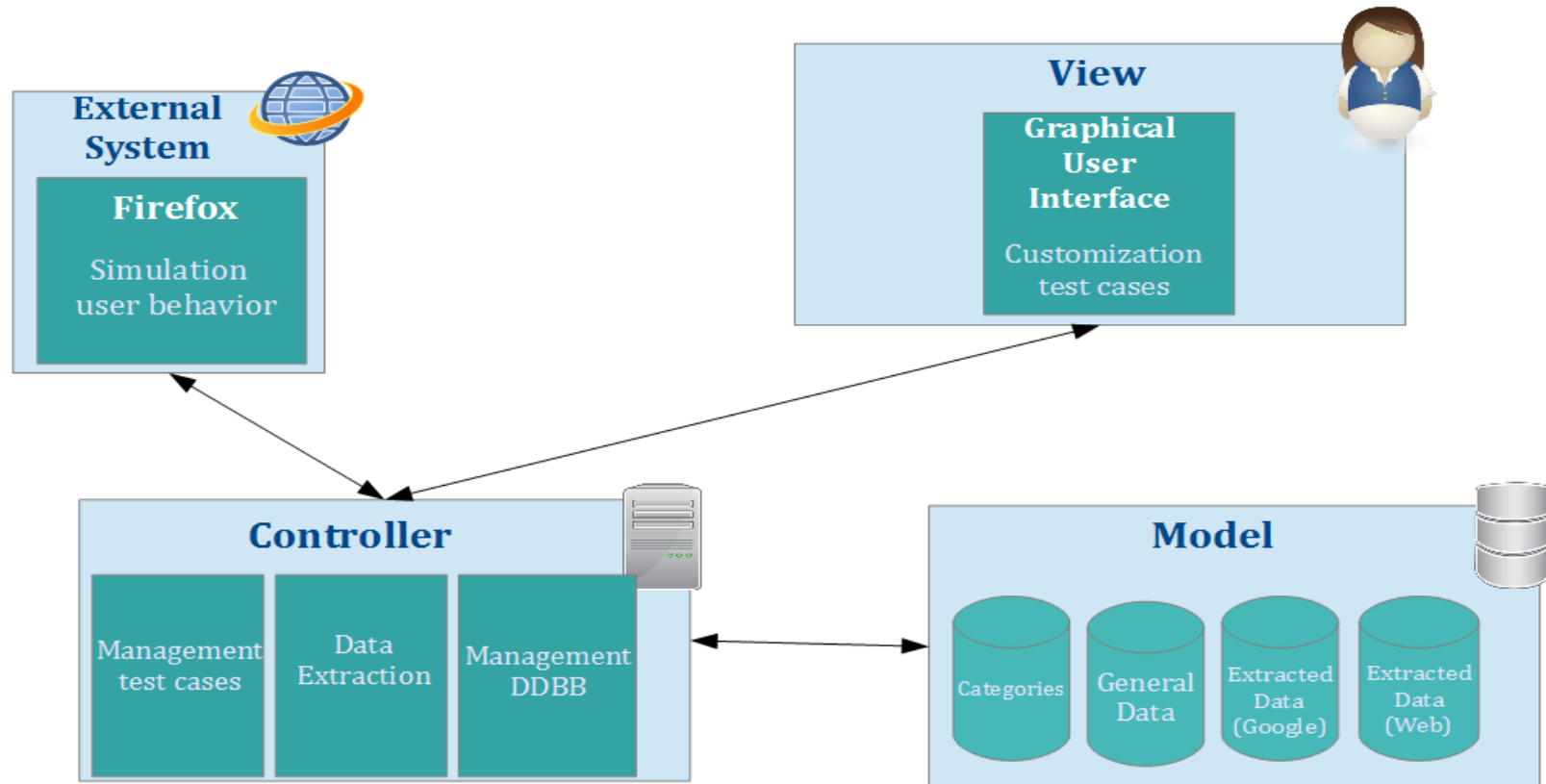


Figure A.3: System Architecture

- **Model:** It's the representation of system information, it also manages all accesses and all changes in such information.

In our case, the model will contain the local database and also the management of requests of the controller.

The database will be divided into 4 sections:

- The first one will store the data about ad interest categories provided by Google, this data won't be modified, only requested by the controller.

- The second one will store the data about the general information of the test cases. It can be modified by the controller.

- The third one will store the relevant extracted data from settings for Google ads -Ads on Google-. It can be modified by the controller.

- The fourth one will store the relevant extracted data from settings for Google ads -Google ads across the web-. It can be modified by the controller.

The chosen database manager will be **MySQL**, that will provide concurrence to the system.

Example of operation: <https://youtu.be/Gb6kjNIDatQ>

Results and Evaluation

The developed tool will simulate the user behavior, using the Google services with 2 sets of words. The evaluation will focus on the data extracted from **Ads on Google**.

- **Set 1:** 50 words that belong to the level 2 from the ad interest categories -provided by Google-.

- **Set 2:** 15 words that belong to the sensitive categories, from the ad interest categories and from the keywords -offered by Google to its advertisers in Google Adwords-. (More information in the following link: <https://support.google.com/adwords/answer/156178?hl=en>).

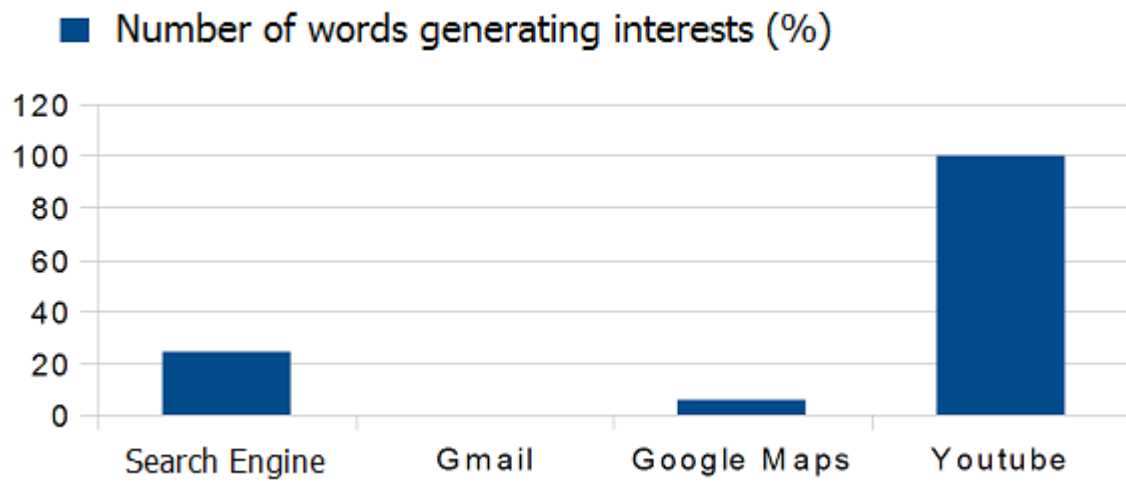


Figure A.4 Number of words generating interests.

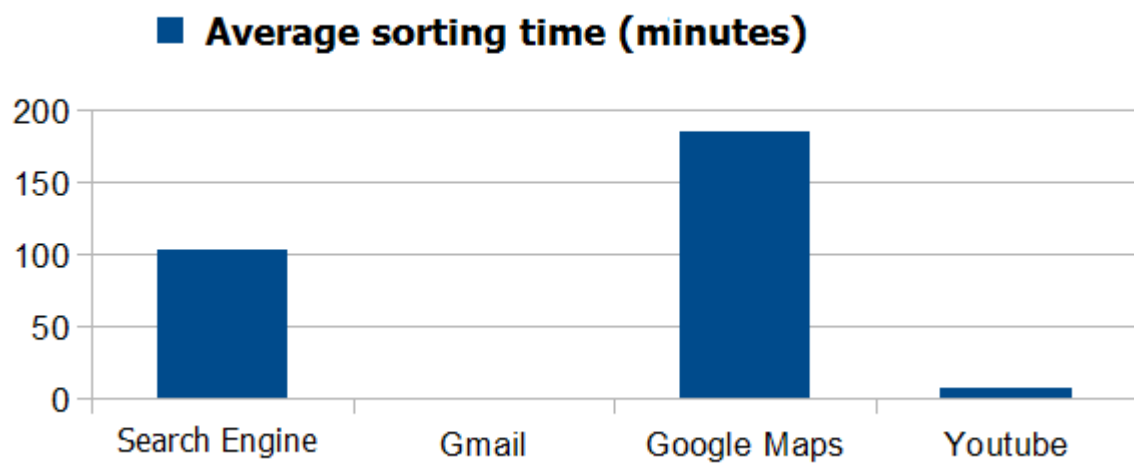


Figure A.5: Average sorting time.

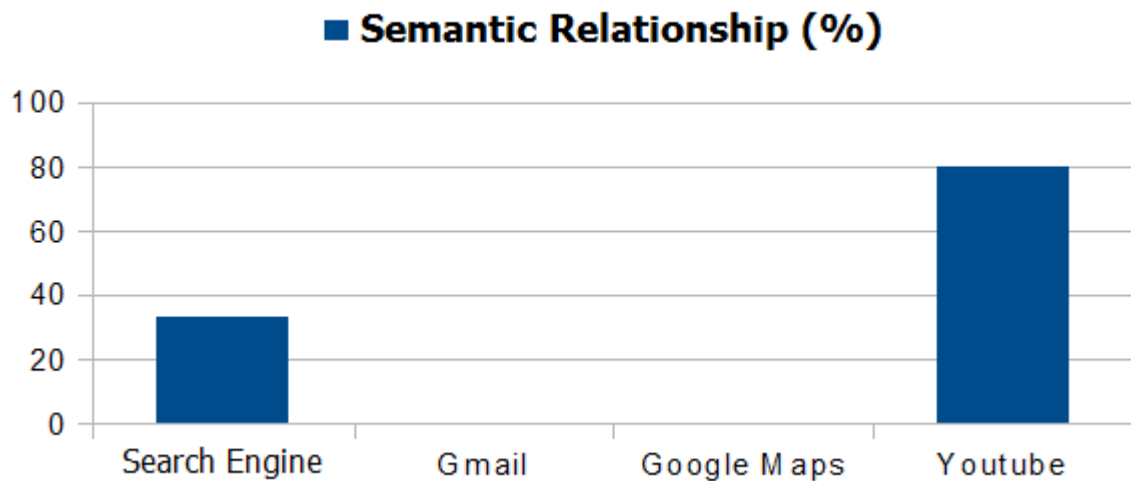


Figure A.6: Semantic Relationship.

Youtube is the service that provides better time performance, more number of words generating new interests and also better semantic relationship. Moreover, it reaches good results only in 5 minutes. Followed by the **search engine** that provides discrete results but, due to its relevance, it's suitable to consider it. Finally, there is not enough information about the use of **Google Maps** and **Gmail** as a marketing tool.

With regards to the **sensitive categories**, it's curious that Google offers keywords and ad interest categories -to its advertisers- that contains some sensitive categories. However, after testing the processing, it hasn't detected that the company doesn't **satisfy** the terms of its privacy policy or the applicable laws on data protection.

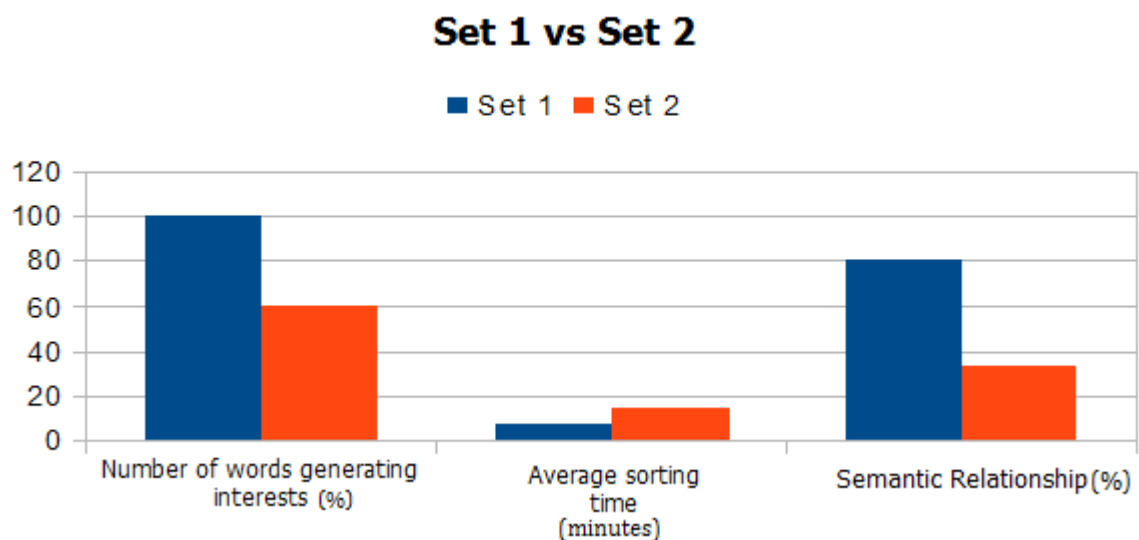


Figure A.7: Set 1 vs Set 2.